PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification : Not classified

A2

(11) International Publication Number: **WO 99/01020**

(43) International Publication Date: 14 January 1999 (14.01.99)

(21) International Application Number: PCT/US98/13608

(22) International Filing Date: 30 June 1998 (30.06.98)

(30) Priority Data:

60/051,480	1 July 1997 (01.07.97)	US
60/051,381	1 July 1997 (01.07.97)	US
60/058,663	12 September 1997 (12.09.97)	US
60/058,598	12 September 1997 (12.09.97)	US

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US).

(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With declaration under Article 17(2)(a); without classification and without abstract; title not checked by the International Searching Authority.

(54) Title: 19 HUMAN SECRETED PROTEINS

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

19 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoeitin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein , a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH_2PO_4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched , for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

Polynucleotides and Polypeptides of the Invention

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

This gene shares sequence homology with The Kruppel family of zinc finger proteins which are thought to be important in embryonic development (See Genebank Accession No. pirlA46017lA46017). Preferred polypeptides comprise the following amino acid sequence:

MSLHVDKEQWMFSICCTACDFVTMEEAEIKTHIGTKHTGED
RKTPSESNSPSSSSLSALSDSANSKDDSDGSQKNKGGNNLLVISVMPGSQPSL
NSEEKPEKGFECVFCNFVCKTKNMFERHLQIHLITRMFECDVCHKFMKTPEQL
LEHKKCHTVPTGGLXXGQW (SEQ ID NO:60);MECHLKTHYKMEYK
CRICQTVKANQL ELETHTREHRLGNHYKCDQCGYLSKTANKLIEHVRVHTG

ERPFHCDQCSYSXKRKDNLNLHKKLKHAPRQTFSCEECLFKTTHPFVFSRHV KKHQSGDCPEEDKKGLCPAPKEPAGPGAPLLVVGSSRNLLSPLSVMSASQALQ TVALSAAHGSSSEPNLALKALAFNGSPLRFDKYRNSDFAHLIPLTMLYPKNHL DLTFHPPRPQTAPPSIPSPKHSFLAYLGLRERAETV (SEQ ID NO:59); and/or LIEHVRVHTGERPFHCDQC (SEQ ID NO:61). Also preferred are the polynucleotides encoding these polypeptides. This gene maps to chromosome 19, and therefore, may be used as a marker in linkage analysis for chromosome 19.

This gene is expressed in several cell types including osteoblasts, T-cells, smooth muscle, and microvascular endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental and immune disorders, including those of the skeletal and muscular systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal, muscular, and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., developing tissue, immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:35 as residues: Ser-30 to Gly-37.

The tissue distribution and homology to Kruppel family of zinc finger proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in T-cells indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer e.g., by boosting immune responses. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and

leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1711 of SEQ ID NO:11, b is an integer of 15 to 1725, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The translation product of this gene was shown to have homology to a *Caenorhabditis elegans* protein (See Genebank Accession No. gil529708). One embodiment of this gene comprises polypeptides of the following amino acid sequence: VDPKKTIQMGSFRINPDGSQ (SEQ ID NO:62), and/or YARSEAHLTELLE (SEQ ID NO:63). An additional embodiment is the polynucleotides encoding these polypeptides.

This gene is expressed primarily in adipose tissue and to a lesser extent in a variety of benign and cancer tissues including tonsils, bladder, placenta spleen, liver cancer, colon cancer, osteosarcoma, chondrosarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer of a variety of tissues and organs, particularly liver, colon, bone and cartlidge. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skelatal, inestinal, reproductive, urinary, and adiplose systems, expression of this

gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., adipose cells or tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:36 as residues: Arg-21 to Leu-26, Arg-88 to Asn-104, Arg-111 to Ser-116, Arg-154 to Lys-160, Cys-164 to Asp-169.

The tissue distribution in tumors of colon, liver, and bone origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of these tumors, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1166 of SEQ ID NO:12, b is an integer of 15 to 1180, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, congenital malformations of the heart. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken

from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue expression within heart tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, treatment, and/or prevention of various disorders of the cardiovascular system. In addition the expression in fetus would suggest a useful role for polynucleotides and polypeptides corresponding to this in developmental abnormalities, fetal deficiencies, pre-natal disorders and various would-healing models and/or tissue trauma. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 895 of SEQ ID NO:13, b is an integer of 15 to 909, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene maps to chromosome 2, and therefore, may be used as a marker in linkage analysis for chromosome 2.

This gene is expressed primarily in infant and adult brain, and placenta and umbilical cord.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, various diseases of the brain, particular mood disorders, and reproductive disorders associated with fetal wasting. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system and female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, and reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine,

synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:38 as residues: Leu-19 to Asn-29, Glu-96 to Gln-107.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses , autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo and/ or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1294 of SEQ ID NO:14, b is an integer of 15 to 1308, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

The translation product of this gene has been shown to have homology to the human GalNAc-T2 gene which is involved in oligosaccaride metabolism/modifications of proteins (See Genebank Accession No. gb|Y10344|HSY10344). This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in fetal heart and to a lesser extent in cerebellum, spleen, thymus, amniotic cells, and fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, Cancers of a variety of tissues, particularly brain, thymus, and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuroendocrine and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, and immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:39 as residues: Ser-19 to His-27, Trp-40 to Ser-45.

The tissue distribution in fetal brain, spleen and thymus tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors of said tissues, in addition to other tumors where expression has been indicated. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Protein, as well as, antibodies directed against the protein may show utility as a tissue-specific marker and/or immunotherapy target for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1970 of SEQ ID NO:15, b is an integer of 15 to 1984, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

The translation product of this gene was shown to have homology to a temperature sensitive supressor in Saccharomyces cerevisiae (See Genebank Accession No. gil987287). According to one embodiment, polpeptides of the invention comprise the sequence:

GCLGFQPPYHSVPAWERSTRGGDHRVELYKVLSSLGYHVVTFDYRGWGDSV GTPSERGMTYDALHVFDWIKARSGDNPVYIWGHSLGTGVATNLVRRLCERET PPDALILESPFTNIREEAKSHPFSVIYRYFPGFDWFFLDPITSSGIKFANDENVKH ISCPLLILHAEDDPVVPFQLGRKLYSIAAPARSFRDFKVQFVPFHSDLGYRHKYI YKS PELPRILREFLGKSEPEHQH (SEQ ID NO:64); YRGWGDSVGTPSERG MTYD (SEQ ID NO:65); and/or ALILESPFTNI (SEQ ID NO:66). Additional embodiments are directed to polynucleotides encoding these polypeptides. This gene maps to chromosome 20, and therefore, may be used as a marker in linkage analysis for chromosome 20.

This gene is expressed is expressed in a broad range of tissues and cell types including lymph node, dendritic cells placenta, monocytes, breast tissue, spleen, brain, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for diagnosis of diseases and conditions which include, but are not limited to, immune disorders including AIDS, autoimmune disorders such as lupus, and respiratory disorders including athsma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, respitory system, and neuroendocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in tonsils indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other

processes that may also suggest a usefulness in the treatment of cancer e.g., by boosting immune responses. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Alternatively, based upon the homology to a known heat shock protein, the translation product of this gene may show utility in normal protein metabolism, including folding, secretion, and proteolytic processing, particularly during periods of increased adrenaline release and stress. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1997 of SEQ ID NO:16, b is an integer of 15 to 2011, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares sequence homology with human growth arrest inducible gene which is a key regulatory molecule in growth stimulation in a variety of tissues. Since such genes may be involved in tumor suppression, the translation product of this gene may be useful in the diagnosis, treatment, and/or prevention of a variety of tumors (See Genebank Accession No.GB:U42437).

This gene is expressed in a variety of tissues including testis, brain, breast, and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disease of the CNS, PNS, and reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, reproductive and respitory systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, and reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:41 as residues: Asp-33 to Lys-41, Arg-109 to Ser-114, Val-127 to Phe-137, Glu-285 to Arg-292.

The tissue distribution and homology to human growth hormone indicates polynucleotides and polypeptides corresponding to this gene are useful for treatment of a variety of diseases, primarily cancers and other proliferative disorders, in which cell growth stimulation is necessary. Alternatively, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses , autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1366 of

SEQ ID NO:17, b is an integer of 15 to 1380, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:17, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

This gene is expressed in kidney, bone marrow, testis, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune, urogenital, or reproductive systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, urogenital, or reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., reproductive tissue, and immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoetic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention

are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2027 of SEQ ID NO:18, b is an integer of 15 to 2041, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

The translation product of this gene shares sequence homology with iduronate sulphate sulphatase (IDS) which is thought to be important for the lysosomal degradation of heparan sulfate and dermatan sulfate. Mutations causing IDS deficiency in humans result in the lysosomal storage of these glycosaminoglycans and Hunter syndrome, an X chromosome-linked disease. This gene maps to the X chromosome, and therefore, may be used as a marker in linkage analysis for the X chromosome.

This gene is expressed primarily in brain, testis, and small intestine.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, Hunter's Syndrome, CNS, skeletal disorders, and/or neural disorders, particularly those associated with abnormalities in lipid and/or oligosaccaride processing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the X-linked disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:43 as residues: Met-1 to Asn-7, Pro-21 to Gly-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimers Disease, Parkinsons Disease, Huntingtons Disease, Hurler's and Hunter's syndrom, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic

disorder, learning disabilities, ALS, psychoses , autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular and skeletal systems. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1861 of SEQ ID NO:19, b is an integer of 15 to 1875, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

The translation product of this gene shares sequence homology with the highly conserved elongation factor G from Rattus norvegicus which is thought to be the protein that promotes the GTP-Dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome in mitochontria (See Genebank Accession No. gil310102). Preferred polypeptides comprise the following amino acid sequence: LDAVLEYLPNPSEVQNYAILNKEDDSKEKTKILMNSSRDNSHPFVGLAFKLEV GRFGQLTYVRSYQGELKKGDTIYNTRTRKKVRLQRLARMHADMMEDVEEVYA GDICALFGIDCASGDTFTDKANSGLSMESIHVPDPVISIAMKPSNKNDLEKFSK GIGRFTREDPTFKVYFDTENKETVISGMGELHLEIYAQRLEREYGCPCITGKPK VAFRETITAPVPFDFTHKKQSGGAGQYGKVIGVLEPLDPEDYTKLEFSDETFGS NIPKQFVPAVEKG FLDACEKGPLSGHKLSGLRFVLQDGAHHMVDSN EIS FIRAGEGALKQALANATLCILEPIMAVEVVAPNEFQGQVIAGINRRHGVITGQD GVEDYFTLYADVPLNDMFGYSTELRSCTEGKGEYTMEYSRYQPCLPSTQE DVINKYLEATGQLPVKKGKAKN (SEQ ID NO:67); SHPFVGLAFKLE (SEQ ID NO:68); RMHADMMEDVEEVYAG DICALFGIDCA SGD (SEQ ID NO:69); LSMESIHVPDPVIS (SEQ ID NO:70), AMKPSNKNDLEKFSKGI (SEQ ID NO:71); RFTREDPTFKV (SEQ ID NO:72); FVLQDGAHHMVDSNEISFIRAGEG ALKQALA

(SEQ ID NO: 73); EDYFTLY ADVPLNDMFGYSTELRSCTEGKGEYTMEY (SEQ ID NO:74); and/or GQLPVKK GKAKN (SEQ ID NO:75). Also preferred are the polynucleotides encoding these polypeptides.

This gene is expressed in many tissues including osteoclasts and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, osteoporosis and prostate cancer, and abnormalities associated with protein metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bones and the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, protstate, skeletal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:44 as residues: Thr-22 to Pro-28.

The homology of this gene to a known translation elongation factor indicates that the gene may show utility in the gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sachs disease, phenylkenonuria, galactosemia, porphyrias, and Hurler's syndrome. Alternatively, expression within osteoclasts may implicate the translation product of this gene as having utility in the detection and treatment of disorders and conditions affecting the skeletal system, in particular the connective tissues (arthritis, trauma, tendonitis, chrondomalacia and inflammation). Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of

a-b, where a is any integer between 1 to 2418 of SEQ ID NO:20, b is an integer of 15 to 2432, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

The translation product of this gene shares sequence homology with thioredoxin which has been demonstrated to be an essential component of the early pregnancy factor activity of serum in pregnant females. In addition, it has been proposed that this gene may be able to confer resistance to specific toxins (i.e. snake venom, etc.). See GenBank No. gil633632). Additional embodiments of this gene are polypeptides comprised of the following amino acid sequences: MGSTVCTDERXMAELAKELPQVSFVKLEAEGVPEVSEKYEISSVPTFLFFKNSQ KIDRLDGAHAPELTKKVQRHASSGSFLPSANEHLKEDLNLRLKKLTHAAPCML FMKGTPQEPRCGFSKQMVEILHKHNIQFSSFDIFSDEEVRQGLKAYSSWPTYPQ LYVSGELIGGLDIIKELEASEELDTICPKAPKLEERLKVLTNKASVMLFMKGNK QEAKCGFSKQILEILNSTGVEYETFDILEDEEVRQGLKAYSNWPTYPQLYVKGE LVGGLDIVKELKENGELLPILRGEN (SEQ ID NO:76); MLFMKGTPQEPRCGFSK QMVEIL (SEQ ID NO:77); and/or WPTYPQLYVSGELIGGLDIIKE (SEQ ID NO:78). Additional embodiments are polynucleotides encoding these polypeptides.

This gene is expressed in placenta, testes, brain, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the reproductive, neural, and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, immune cells and tissue, and reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:45 as residues: Leu-15 to Asp-20.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in bone marrow combined with its homology to thioredoxin, indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer e.g., by boosting immune responses. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Alternatively, the tissue distribution may suggest that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1255 of

SEQ ID NO:21, b is an integer of 15 to 1269, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

Other embodiments of the gene include polypeptides comprised of the following amino acid sequences: FKHRGLEYGRFLRXWELKPEFXKGFRTDGRAGXWVXGDFGKRFFRPGEVAD SCNPSTFGXRGWQITCRPGV (SEQ ID NO:79); GDFGKRFFRPGEVADSCNPST FG (SEQ ID NO:80); MGGQVXGSXXILEKDFSGQVRWLIPVIPALLEXEAGRSL VGQ EFETSLGNMAKPCLYKNYKISARSGGLCL (SEQ ID NO:81); ILEKDFSG QVRWLIP VIPALLE (SEQ ID NO:82); and EAGRSLVGQEFETSLGNMAKPC LYKNYK ISARSGGLCL (SEQ ID NO:83). Additional embodiments include polynucleotides encoding these polypeptide sequences.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the brain, such as Alzheimer's and Parkinson's diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., neural tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the

developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:22 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 748 of SEQ ID NO:22, b is an integer of 15 to 762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:22, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

The translation product of this gene shares sequence homology with olfactomedian which is thought to be an important component in the extra cellular matrix of the neuroepithelium. By analogy to other extracellular matrix proteins of the nervous system, olfactomedin may influence the maintenance, growth, or differentiation of chemosensory cilia on the apical dendrites of olfactory neurons. Other embodiments of this gene include polypeptides comprised of the following amino acid sequences:

MTVGPASALFPCQTPXFPWTEWNXWEFTAHVLSQKFEKELSKVREYVQLISVY EKKLLNLTVRIDIMEKDTISYXELDFELIKVEVKEMEKLVIQLKEPFGGSSEIVGP AGGGDKKYDSLGREA (SEQ ID NO:84), MTLLVEKLETLDKNXVLAIRREXVAL KTKLKECEASKDQNTPVVHPPPTPGSCGHGGVVXISKPSVVQLNWRGFSYLY GAWGRDYSPQHPNKGLYWVAPLNTDGRLLEYYRLYNTLDDLLLYINARELRIT YGQGSGTAVYNNNMYVNMYNTGNIARVNLTTNTIAVTQTLPNAAYNNRFXY ANVAWQDIDFXVDENGLWVIYSTEASTGNMVISKLNDTTLQVLNTWYTXQYK PSASNAFMVCGVLYATRTMNTRTEEIFYYYDTNTGKEGKLDIVMHKMQEKVQ SINYNPFDQKLYVYNDGYLLNYDLSVLQKPQ (SEQ ID NO:85), LETLDKNX VLAIRREXVALKTKL KECE (SEQ ID NO:86), YWVAPLNTDGRLLE (SEQ ID NO:87), ASNAFMVCGVLY (SEQ ID NO:88), and/or TGKEGKLDIVM (SEQ ID NO:89). Additional embodiments are polynucleotides encoding these polypeptides. This gene maps to chromosome 13, and therefore, may be used as a marker in linkage analysis for chromosome 13.

This gene is expressed primarily in small intestine and pancreas, also during ulcerative colitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the digestive tract, such as inflammatory bowel disease, and pancreatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, especially the small intestine and pancreas, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., gastrointestinal tissue, digestive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., bile, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:47 as residues: Ser-48 to Ser-59, Val-77 to Cys-83.

The homology to a known protein thought to be involved in the maintenance, growth, and/or differentiation of chemosensory cilia on the apical dendrites of nuerons indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioural disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, protein may show utility in the diagnosis, treatment, and/or prevention of various olfactory and sensory disorders. Alternatively,the tissue distribution in gastrointestinal tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sachs disease, phenylkenonuria, galactosemia, porphyrias, and Hurler's syndrome.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2874 of SEQ ID NO:23, b is an integer of 15 to 2888, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

The translation product of this gene shares sequence homology with aspartyl beta-hydroxylase. Aspartyl beta-hydroxylase specifically hydroxylates a single Asp or Asn residue in certain epidermal growth factor-like domains of a number of proteins and thus may play a major role in the differentiation and development of cells (See GenBank No.il162694). One embodiment of this gene comprises polypeptides of the following amino acid sequence: MSRLLAKAKDFRYNLSEVLQGKLGIYDADGDGDFDVDDAK VLLGLTKDGSN ENIDSLEEVLNILAEESSDWFYGFLSFLYDIM TPFEMLEEEEEE SETADGVDGT SQNEGVQGKTCVILDLHNQ (SEQ ID NO:90), TSAGSSSPGTRER DKAWRTQQ WEERRTLRNFILHVVYGDCIAGRLDICTCRLV (SEQ ID NO:91), RVRAAAAPAR GRETKHGGHNN (SEQ ID NO:92), and/or SFFTWFMVI ALLGVWTSV (SEQ ID NO:93). An additional embodiment are polynucleotides encoding these polypeptides. This gene maps to chromosome 8, and therefore, may be used as a marker in linkage analysis for chromosome 8.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the brain and central nervous system, such as Alzheimer's and Parkinson's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g.,

neural tissue, differentiating tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:48 as residues: Ile-40 to Lys-45.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, learning disabilities, ALS, psychoses, autism, and altered bahaviors, including disorders in feeding, sleep patterns, balance, and preception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Alternatively, the homology to a conserved protein that specifically modifies signal transduction proteins may suggest that the protein is beneficial in the diagnosis, treatment, and/or prevention of various disorders affecting proliferating tissues, such as as cancer. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1368 of SEQ ID NO:24, b is an integer of 15 to 1382, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

Additional embodiments of this gene include polypeptides comprised of the following amino acid sequences:

WCQRVQDLSARVRGEQCCAVGRNLTITQSPRQRVQDLSTGVRGEQRCPAGRSL
TITQSPHRHPVSSPEGPGPQCRGARRAVLSSGEEPHHHSVSSPAHFFSMSRFAP

PLVFVFLKEDFEKRW (SEQ ID NO:94); NQLTFIWKKPHFTVVCHFDGVRGSRT SVPG CEESSAVQWGGTSPSPSLLARGSRTSVPGCEESSAVQRGGVSPSPSLLTV TQSPRQRVQDLSAGVRGEQCCPAGRNLTITQSPHQHTFSPCLVLLLLWYLYFLK RILKRDGEVGILGRRDQLFPQD (SEQ ID NO:95); LSFGKSPTSLWSVTLM VSEGPGPQCQGARRAVLCSGEEPHHHPVSSPEGPGPQYRGARRAALSSGEESH HHPVSSPSPSLLARGSRTSVPGCEESSAVQRGGTSPSLSLLTSTLFLHVSFCSSS GICIS (SEQ ID NO:96); and MVSEGPGPQCQGARRAVLCSGEEPHHHPVS SPEGPGPQYRG ARRAALSSGEESHHHPVSSPSPSLLARGSRTSVPGCEESSA VQRGGTSPSLSLLTSTLFLHVSFCSSSGICIS (SEQ ID NO:97). Additional embodiments include polynucleotides encoding these polypeptides.

This gene is expressed primarily in human adrenal gland tumor and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, carcinoma, reproductive, and/or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., endocrine tissue, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of endocrine disorders and cancers (e.g., Addison's disease, Cushing's syndrome, Thyrotoxicosis, metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in placenta would suggest that polynucleotides and polypeptides corresponding to this gene are useful in diagnostics and therapeutics relating to developmental abnormalities, fetal deficiencies, pre-natal disorders and would-healing and/or tissue traumas. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related

polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1642 of SEQ ID NO:25, b is an integer of 15 to 1656, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

The translation product of this gene shares sequence homology with ATP-dependent RNA helicase which is thought to be important in gene transcription (See GenBank No. gil914885). One embodiment of this gene comprises polypeptides of the following amino acid sequence: GLCTEVAFAASLRGPSAHIISDPQTTLQRGGRCCKLHSSPNWHHPASWDSDQG CQTPEPVVLSLHLSARPPPWSGFLSFLLQVSFSLCYHLCSEQLLTTQRVSCAHIY SALDPTARKINLAKFTLGKCSTLIVTDLAARGLDIPLLDNVINYSFPAKGKLFLH RVGKQPVAGPGAGRGAGSWQKPRVQGLTLDTAHGVAVGLVLETEPRYIA (SEQ ID NO:98), GIEKFGNLPKVTQLVCSRIRIR LVH (SEQ ID NO:100); KSLVT CPRSHSLFVAESG (SEQ ID NO:101); VFHVETLFSALYILTHVILIIRHKEGAVIRT DEENEA (SEQ ID NO:102); and/or VTDLAARGLDIPLLDNVINYSF (SEQ ID NO:99). Additional embodiment are polynucleotides encoding these polypeptides.

This gene is expressed primarily in B-cell lymphoma and neutrophil.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, lymphoma and other immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in B-cells combined with the homology to an RNA-dependent helicase indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer e.g., by boosting immune responses. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1137 of SEQ ID NO:26, b is an integer of 15 to 1151, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

Additional embodiments of the gene include polypeptides comprised of the following amino acid sequences:

TFQFCHTHQPCTCPSHHSGYKSISLWFWLCPNDCEAEHLFKCELAIYIPSLENC LFKPFAPFYIELSIF (SEQ ID NO:103); LYYFIFPPAVNKHSNFAILTNLVLVQAII VGIKVFPCGSGYALMTVRLNIFSSVNWPFIYLLWRTVFSNPLLLFTLSYPSFNC WVVYCLI (SEQ ID NO:104);

This gene is expressed primarily in human bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematapoiesis and leukemias. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoetic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1285 of SEQ ID NO:27, b is an integer of 15 to 1299, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in jurkat cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T-cell related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., immune cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Expression of this gene product in Jurket cells indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product may be involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer e.g., by boosting immune responses. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the

present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 857 of SEQ ID NO:28, b is an integer of 15 to 871, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where the b is greater than or equal to a + 14.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element) pathway. Thus, it is likely that this gene activates T-cells through the Jaks-STAT signal transduction pathway. The ISRE is a promoter element found upstream in many genes which are involved in the Jaks-STAT pathway. The Jaks-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. This gene maps to chromosome 3, and therefore, may be used as a marker in linkage analysis for chromosome 3. Additional embodiments of the invention are directed to polypeptides comprising the following amino acid sequences: HQAPTQSQLGNQSHPPWLCWGGPAICPWSRRERGVSPRPGAGKECVPQLSAL LILIMEKPLFLSPFPELVFCCFCFILFWGDSFLLFNLESPVPLGCRQFLPGPSRNP HSPSPLLRYLQEAANLVHSDKPPTQISLLPLCPKSHH (SEQ ID NO:105) and MEKPLFL SPFPELVFCCFCFILFWGDSFLLFNLESPVPLGCRQFLPGP SRNPHSPSPLLRYLQEAANLVHSDKPPTQISLLPLCPKSHH (SEQ ID NO:106). Further embodiments are directed to polynucleotides encoding these polypeptides.

This gene is expressed primarily in human gall bladder.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic and gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., hepatic tissue, pancreatic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., bile, serum, plasma, urine, synovial fluid and spinal fluid) or another

tissue or cell sample taken from anindividual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in gall bladder indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sachs disease, phenylkenonuria, galactosemia, porphyrias, and Hurler's syndrome. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells and in lipid metabolism). In addition the expression in fetus would suggest a useful role for polynucleotides and polypeptides corresponding to this gene in developmental abnormalities, fetal deficiencies, pre-natal disorders and various would-healing models and/or tissue trauma. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1009 of SEQ ID NO:29, b is an integer of 15 to 1023, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where the b is greater than or equal to a + 14.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3'NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HTXBK30	209118 06/05/97	Uni-ZAP XR	11	1725	1	1493	10	10	35	1	22	23	55
2	H2MBB56	209118 06/05/97	pBluescript SK-	12	1180	457	1180	513	513	36	1	22	23	182
3	HHFHV86	209118 06/05/97	Uni-ZAP XR	13	909	1	909	141	141	37	1	20	21	21
4	HIBCW32	209118 06/05/97	Other	14	1308	36	1308	986	986	38	1	53	54	107
4	HIBCW32	209118 06/05/97	Other	31	1361	144	1361	203	203	55	1	18	19	60
5	HLHCI58	209118 06/05/97	Uni-ZAP XR	15	1984	462	1984	585	585	39	1	31	32	47
6	HLMFG37	209118 06/05/97	Lambda ZAP II	16	2011	1095	2011	1345	1345	40	1	21	22	82
7	HTLFA90	209118 06/05/97	Uni-ZAP XR	17	1380	38	1364	181	181	41	1	32	33	314
8	HRDDR94	209118 06/05/97	Uni-ZAP XR	18	2041	247	1101	1452	1452	42	1	48	49	82
8	HRDDR94	209118 06/05/97	Uni-ZAP XR	32	1822	942	1822	1095	1095	56	1	18	19	39
9	HSIDY06	209118 06/05/97	Uni-ZAP XR	19	1875	1	1875	1471	1471	43	1	25	26	45
9	HSIDY06	209118 06/05/97	Uni-ZAP XR	33	1873	1	1873		275	57	1	18	19	27
10	HSKGO49	209118 06/05/97	pBluescript	20	2432	125	1613	296	296	44	1	31	32	45

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3'NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
11	HAGDU63	209118 06/05/97	Uni-ZAP XR	21	1269	18	1245	266	266	45	1	20	21	20
12	HBXGM67	209118 06/05/97	ZAP Express	22	762	1	762	32	32	46	1	33	34	42
13	HUFAC36	209118 06/05/97	pSport1	23	2888	1	2888	19	19	47	1	21	22	113
14	HAGBZ81	209118 06/05/97	Uni-ZAP XR	24	1382	24	1382		65	48	1	30	31	49
15	HATCI19	209118 06/05/97	Uni-ZAP XR	25	1656	1	1656	77	77	49	1	29	30	37
16	HBJCK69	209118 06/05/97	Uni-ZAP XR	26	1151	1	1151	56	56	50	1	24	25	45
17	HBMDD55	209118 06/05/97	pBluescript	27	1299	1	1299	111	111	51	1	17	18	25
18	HCACJ81	209118 06/05/97	Uni-ZAP XR	28	871	1	871	191	191	52	1	24	25	49
18	HCACJ81	209118 06/05/97	Uni-ZAP XR	34	865	1	865	460	460	58	1	55	56	69
19	HCE3F11	209118 06/05/97	Uni-ZAP XR	29	1023	1	1018		173	53	1			10

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly , the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification , such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown inTable 1, the ORF (open reading frame), or any fragement specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the presence invention can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identiy are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the lenght of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignement of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determing the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is becuase the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequnce are manually corrected for. No other manual corrections are to made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E. coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-

60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein

molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein

by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiollitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox , hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteocarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stoke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotaxic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotaxic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotaxic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors),or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, caricadic rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lafmid BA	plafmid BA
pSport1	pSport1
pCMVSport 2.0	pCMVSport 2.0
pCMVSport 3.0	pCMVSport 3.0
pCR®2.1	pCR®2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 µl of reaction mixture with 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM $MgCl_2$, 0.01% (w/v) gelatin, 20 µM each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P^{32} using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D.^{600}$) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

The following alternative method can be used to purify a polypeptide expressed in *E coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfuidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 μm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., Virology 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five μg of a plasmid containing the polynucleotide is co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One μg of BaculoGold™ virus DNA and 5 μg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ^{35}S-methionine and 5 µCi ^{35}S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from

Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No.209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five µg of the expression plasmid pC6 is cotransfected with 0.5 µg of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of metothrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 µM, 2 µM, 5 µM, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 µM. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the

polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCCAAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA

GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with

this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl2 (anhyd); 0.00130 mg/L $CuSO_4$-$5H_2O$; 0.050 mg/L of $Fe(NO_3)_3$-$9H_2O$; 0.417 mg/L of $FeSO_4$-$7H_2O$; 311.80 mg/L of Kcl; 28.64 mg/L of $MgCl_2$; 48.84 mg/L of $MgSO_4$; 6995.50 mg/L of NaCl; 2400.0 mg/L of $NaHCO_3$; 62.50 mg/L of NaH_2PO_4-H_2O; 71.02 mg/L of Na_2HPO4; .4320 mg/L of $ZnSO_4$-$7H_2O$; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H_2O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H_2O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H_2O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-

Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalainine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tryrosine-2Na-2$H_2$0; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B_{12}; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proxial region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

Ligand	JAKs tyk2	Jak1	Jak2	Jak3	STATS	GAS(elements) or ISRE
IFN family						
IFN-a/B	+	+	-	-	1,2,3	ISRE
IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
Il-10	+	?	?	-	1,3	
gp130 family						
IL-6 (Pleiotrohic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
Il-11(Pleiotrohic)	?	+	?	?	1,3	
OnM(Pleiotrohic)	?	+	+	?	1,3	
LIF(Pleiotrohic)	?	+	+	?	1,3	
CNTF(Pleiotrohic)	-/+	+	+	?	1,3	
G-CSF(Pleiotrohic)	?	+	?	?	1,3	
IL-12(Pleiotrohic)	+	-	+	+	1,3	
g-C family						
IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
IL-7 (lymphocytes)	-	+	-	+	5	GAS
IL-9 (lymphocytes)	-	+	-	+	5	GAS
IL-13 (lymphocyte)	-	+	?	?	6	GAS
IL-15	?	+	?	+	5	GAS
gp140 family						
IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
IL-5 (myeloid)	-	-	+	-	5	GAS
GM-CSF (myeloid)	-	-	+	-	5	GAS
Growth hormone family						
GH	?	-	+	-	5	
PRL	?	+/-	+	-	1,3,5	
EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
Receptor Tyrosine Kinases						
EGF	?	+	+	-	1,3	GAS (IRF1)
PDGF	?	+	+	-	1,3	
CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:
5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCG
AAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:
5':CTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAATG
ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, Il-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml genticin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1 x 10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest $2x10e^7$ U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM $Na_2HPO_4.7H_2O$, 1 mM $MgCl_2$, and 675 uM $CaCl_2$. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting $1x10^8$ cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of $5x10^5$ cells/ml. Plate 200 ul cells per well in the 96-well plate (or $1x10^5$ cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon

activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

(Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as $5x10^5$ cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to $1x10^5$ cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF-κB (Nuclear Factor κB) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF-κB regulates the expression of genes involved in immune cell activation, control of apoptosis (NF-κB appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κB is retained in the cytoplasm with I-κB (Inhibitor κB). However, upon stimulation, I- κB is phosphorylated and degraded, causing NF- κB to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κB include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF-κB promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF-kB would be useful in treating diseases. For example, inhibitors of NF-κB could be used to treat those diseases related to the acute or chronic activation of NF-kB, such as rheumatoid arthritis.

To construct a vector containing the NF-κB promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF-κB binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:

5' :GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5' :GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5' :CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGACTTTCC ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCA TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTC CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT: 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF-κB/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF-κB/SV40/SEAP cassette is removed from the above NF-κB/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF-κB/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF-κB/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described

in Example 13. As a positive control, exogenous TNF alpha (0.1,1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 µl of 2.5x dilution buffer into Optiplates containing 35 µl of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 µl Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 µl Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25

28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO_2 incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is

incubated at 37°C in a CO_2 incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10^6 cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10^6 cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca^{++} concentration.

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating

tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boeheringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and

PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg_{2+} (5mM ATP/50mM $MgCl_2$), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM $MgCl_2$, 5 mM $MnCl_2$, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavadin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phospotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR

products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or mirocapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2- hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression

of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial

space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The

template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

(1) GENERAL INFORMATION:

(i) APPLICANT: Human Genome Sciences, Inc. et al.

(ii) TITLE OF INVENTION: 19 Human Secreted Proteins

(iii) NUMBER OF SEQUENCES: 106

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

(B) COMPUTER: HP Vectra 486/33

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE: **June 30, 1998**

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: A. Anders Brookes

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PZ009PCT

(vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 733 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCCCAAATCT TCTGACAAAA CTCACACATG CCCACCGTGC CCAGCACCTG 60

AATTCGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCCAAA ACCCAAGGAC ACCCTCATGA 120

TCTCCCGGAC TCCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG 180

TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG 240

AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT 300

GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG 360

AGAAAACCAT CTCCAAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC 420

CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT 480

ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC AACTACAAGA 540

CCACGCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG 600

ACAAGAGCAG GTGGCAGCAG GGGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC 660

ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC 720

GACTCTAGAG GAT 733

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCCCCG AAATGATTTC 60

CCCGAAATAT CTGCCATCTC AATTAG 86

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATTT CCCCGAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTTCCCCG 60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120

GCCCCTAACT CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT 180

TTATGCAGAG GCCGAGGCCG CCTCGGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240

TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGGACTTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCGGCCTCGA GGGGACTTTC CCGGGGACTT TCCGGGGACT TTCCGGGACT TTCCATCCTG 60

CCATCTCAAT TAG 73

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60

CAATTAGTCA GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 120

CAGTTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180

GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240

CTTTTGCAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1725 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

AAGGCAGTGA TGGGAAGAAA CATCCTTATT ATTACAGTTG TCACGTGTGT GGATTTGAGA 60

CCGAGCTCAA TGTCCAGTTT GTCAGCCACA TGTCACTCCA CGTGGACAAG GAGCAGTGGA 120

TGTTTTCRAT CTGCTGCACT GCCTGCGACT TCGTCACCAT GGAGGAAGCA GAGATAAAGA 180

CTCACATTGG CACCAAGCAC ACAGGGGAAG ACAGGAAGAC CCCCAGCGAA TCAAATAGCC 240

CCTCTTCATC CTCCCTCTCA GCTCTGAGTG ATTCAGCCAA CAGCAAAGAT GATTCAGATG 300

GCTCCCAGAA AAACAAGGGC GGGAACAATC TGCTGGTCAT CTCTGTCATG CCTGGGAGCC 360

AGCCCTCACT GAACAGTGAG GAAAAGCCAG AGAAAGGGTT CGAATGTGTT TTTTGCAACT 420

TTGTCTGCAA GACGAAGAAC ATGTTTGAGC GTCATCTGCA GATACACCTC ATCACCCGGA 480

TGTTTGAGTG TGATGTGTGC CACAAGTTCA TGAAGACCCC CGAACAGCTG CTGGAGCATA 540

AGAAATGCCA CACTGTCCCC ACCGGTGGGC TCAASTMAGG ACAGTGGTGA GTTTCAGACT 600

CCTCTAGGTG CCCATTCTGC ATTTATTCCA CCAACCGCCC CGCTGCCATG GAGTGCCACC 660

TCAAGACCCA CTACAAGATG GAGTACAAGT GCCGGATCTG CCAGACGGTG AAGGCCAACC 720

AGCTGGAGCT GGAGACGCAC ACCCGGGAGC ACCGCCTGGG CAACCACTAC AAGTGCGACC 780

AGTGCGGCTA CCTGTCCAAG ACCGCCAACA AGCTCATCGA GCACGTGCGC GTCCACACCG 840

GGGAGCGGCC CTTCCACTGT GACCAGTGCA GCTACAGCTG MAAGCGCAAG GACAATCTCA 900

ACCTGCACAA GAAGCTGAAG CACGCCCCAC GCCAGACCTT CAGCTGCGAA GAGTGCCTGT 960

TCAAGACCAC ACACCCTTTC GTYTTCAGCC GCCACGTCAA GAAGCACCAG AGTGGGGACT 1020

GCCCCGAGGA GGACAAGAAG GGCCTGTGTC CAGCCCCCAA GGAACCGGCC GGCCCGGGGG 1080

CCCCGCTCCT GGTGGTCGGG AGCTCCCGGA ATCTCCTGTC TCCCCTGTCA GTTATGTCTG 1140

CCTCCCAGGC TCTGCAGACC GTGGCCCTGT CGGCAGCCCA CGGCAGCAGC TCAGAGCCCA 1200

ACCTGGCACT CAAGGCTTTG GCCTTCAACG GCTCCCCTTT GCGCTTTGAC AAGTACCGGA 1260

ACTCAGATTT TGCCCATCTC ATTCCCTTGA CAATGTTATA CCCCAAGAAC CACTTGGATC 1320

TCACATTCCA CCCTCCCCGA CCTCAGACTG CGCCTCCCAG CATCCCCTCA CCCAAACACT 1380

CCTTCCTGGC CTATCTCGGA CTGAGAGAAA GAGCAGAGAC TGTCTGAGGG CAGCCATGTT 1440

CTGTACCAAA AACAGAGAGA CAAAAGACAA AAAAAAAAAA AAAAACCACA AAACTTAAAC 1500

ACAACCCCAG CAGGTGTATG TTGCTGCAAA ACCTACAGAC CCCGATGGGT CTGGGAACAT 1560

GTGTACTGTA TATCCTTTCA GTAAGGAATA GAAAATTGGC TCTCGGGTGG TATACCTATT 1620

NGCATTGGAC CTGGAAAGCT GGCTTTTTAT CCAATCTTTC AAGAGAGGTG ACCCTACTGG 1680

CATACTTTCT ANCTTCAGAG GCATGGCTCC CCCAGGCNAC CCAAG 1725

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1180 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

CCGTTTTGAA GGTCCTAGCC CACCTGGTNN GNCTCACGCG CACGACTAGC CGCTCCCATA 60

CAGCACGCCC GGACTCTGTC GTCGCTTAAG GCCACTCCTA TTCTACGGCT GACCCCTGGT 120

GGTCACGTGG ATCTGTTCGC CACGCAAGTC TGGGTCCTTC GGCGATTGAC CGGGGTCCTT 180

GCTGTTCGGG AGCCTCTCCT AAGCTGCCTG TTCGCGCGAR AKTTTGGAGG GGCGGGTTTG 240

GGGTCGGTGT CTGATTGGGG CTCGCACCGC AGCACGCTGG AGTCCCGCTT AGGTACCAGT 300

TAGCGTCAGG GGAGCTGGGT CAGGCGGTCG CCGGGACACC CCGTGTGTGG CAGGCGGCGA 360

AGCTCTGGAG AATCCCGGAC AGCCCTGCTC CCTGCAGCCA GGTGTAGTTT CGGGAGCCAC 420

TGGGGCCAAA GTGAGAGTCC AGCGGTCTTC CAGCGCTTGG GCCACGGCGG CGGCCCTGGG 480

AGCAGAGGTG GAGCGACCCC ATTACGCTAA AGATGAAAGG CTGGGGTTGG CTGGCCCTGC 540

TTCTGGGGGC CCTGCTGGGA ACCGCCTGGG CTCGGAGGAG CCAGGATCTC CACTGTGGAG 600

CATGCAGGGC TCTGGTGGAT GAACTAGAAT GGGAAATTGC CCAGGTGGAC CCCAAGAAGA 660

CCATTCAGAT GGGATCTTTC CGGATCAATC CAGATGGCAG CCAGTCAGTG GTGGAGGTGC 720

CTTATGCCCG CTCAGAGGCC CACCTCACAG AGCTGCTGGA GGAGATATGT GACCGGATGA 780

AGGAGTATGG GGAACAGATT GATCCTTCCA CCCATCGCAA GAACTACGTA CGTGTAGTGG 840

GCCGGAATGG AGAATCCAGT GAACTGGACC TACAAGGCAT CCGAATCGAC TCAGATATTA 900

GCGGCACCCT CAAGTTTGCG TGTGAGAGCA TTGTGGAGGA ATACGAGGAT GAACTCATTG 960

AATTCTTTTC CCGAGAGGCT GACAATGTTA AAGACAAACT TTGCAGTAAG CGAACAGATC 1020

TTTGTGACCA TGCCCTGCAC ATATCGCATG ATGAGCTATG AACCACTGGA GCAGCCCACA 1080

CTGGCTTGAT GGATCACCCC CAGGAGGGGA AAATGGTGGC AATGCCTTTT ATATATTATG 1140

TTTTTACTGA AATTAACTGA AAAAATATGA AACCAAAAGT 1180

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 909 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GTTTTGAAGT AAAATTGACC AGCCAAATTT ATAGGTAGTC TGCACAATTT TGTATCCTTT 60

TTTAATAATG AAAAATTACT ATGAAGAAAT ACTGAACAAA TTTTTATGTG CAATATTTTA 120

TAGACCTATG TATCTGAAGC ATGTTTACAC TGGCGTTTTT TTTTTTAATT AATTTCCTAA 180

ATGTTAAGTA TGATAGAMCA AGCTGACCCA AATCCTTAAG TTTACAAAGC TGTTGGAAAA 240

CTTTGTGTCC TGATTTCAAC AATCACGYTT TGTTTGAAAG ATGAGCCAAG CTCACAGACA 300

CTAAATTTTA TGTCATGCCA TAAGCTGGAG AGGAGCCATT TGGCTACAGC TGCGGAACTT 360

CATTGAGGAG CAAATGAAAG GCACATGGTA CGAGCACGCT GGTGCAGTTC ATGTTCTTCC 420

TGCCTGTGAA TTGAATACTG TCCTGGTAGC AGTTTTGGGT CGGTCAGGAG CTCAAGGCTG 480

GTTTGTGTGG CTGACTACGG ATGAGCACTG AAGTTGCCTC AAAGAATTAA KGGGTGTCCA 540

CACCAGCCTC TGGGGGTCTT TGGTGTTAGT CTTCCAGGTA GAGCTGGTTT TACAAGTAGG 600

TGGCCATCTA CAGGATGTGA TGTGAGCGAT GCCAGACAGC TCTCTCTGAC CCCAGGTAAT 660

GCCCTGAATC TGGTGATCCT GGCTGATCTG TGACCAATAG AGATTAGCTC CTTGGGATTT 720

GGGGTCCTAA AAGGTCCCTG AAAAAATGCA CCCCTTGTCT TTAAGCCAAC ATTGGTGAAG 780

GAACTGAGAA CTCTTAGGGT TACATAAARA AAGACCCCTG TTGAGATAGT TTATGCAGAT 840

ACYTGGRAGG AARTAGAGAT GCCAGGAGGA ATTCTGAGAC CGGCCATAGA GCGNGNCANA 900

AGGCTCCAA 909

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1308 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

AATCCTCAGT CTGGTCTATA TTGTTACTTG CCTGTNTCCA CTAGAAGGAG TCACCCAGGC 60

ATTTAGTGAA TGATAATCCT AGAAGTTGAG TCTGAGTACA TCAGAGTTTG GCATTATTTG 120

CAACATCCCA GGAATGAGCT ATAAGACTCC ATTGCTTTCT GGACTTTAGT GACCAGCTGA 180

TCTTCCCTTG GATCCTCCAA TAAAGGGGAA AAAATACTTG TCTTATGTTG TTTTAGCAAG 240

GGAGTTCAGC CAAGTGACAG GAAAGATTCC CTGGTCTGCA TCTAAACATT AGTCAACAGA 300

AGGAAAAGGC TCTTTTCCTT GGAGATTTTC AAGTTGAAAA AGTACAGCTG CTTGGATGGA 360

GTTTCTCTCA TCAGAATAGA AATATGCTAG GTGGTCTTCA AGACCTTTTT GAACCTACAG 420

ATTCTAAACC TTAGAGGAAG CCAGAATATG TGATCATACC AGGTGAGGAA AGTTAGGAGA 480

TATTCATCTT AAGAGATATC CTATTTGGCA GTCACATGTT ACACTTGAGC AACAATTGTC 540

TAGGCTAGAA ATATAGAACC ACACAATATT TATCATCATT GGGTCAATTT CTCCCCCTCT 600

ATCAAGTGAG GAGATTAAGG CTTAAAGAAA GGAAAGGACT TGCCAGCCAC CACAGGGCTC 660

ATGAGGGGCT TGAACCCATG TCGTGTGTCT CTACCACAGC TGACTCTCAG CTGGCTAATG 720

GAGAAAATGT GAGGAATTAC CTGTCAAATT AGATTGAGCT CAGAGTAAAT AGATGAGCAC 780

ATTTATAACT CTAAATTAGA CTTTCTATCA AATGGGACCA AACATATGGA AGGGGCTGCT 840

GTCCTGCCTC AATTTAGCCA AGACTTGTCT TGTCTCAAAG CCAAGACTCA GCAGTACATG 900

GATTCTGAGA AGTCCAGGGT CTGAATTGTT GTCTCTGATT ACTGGAAGGA CAGGTTAACT 960

GAATGTCTGT GATCACTAAC AGGTGATGGG CTTTGTGCCC ACTCCAGAGA TATTGTGGGA 1020

GACAAATTCT TTTAACAGCC TGTCCTCCCG GCATCAGGAG TCATTGAACA ATCATGGATT 1080

GTTGTGTTTG GGATTTTTTT TTTTTTGGC TTTGTTTTTG GTTTTTGTGT GTGTGTGTGT 1140

GTGCTGCATG CACATGCGTC CAAGACTTAC GGGTGGTTTA GGGGAAAGTG CTAGCAACAG 1200

TGTCAATGTC ATTGTTAACT ATATTTCATA TTTGAGGTCT CATTGTTTTC AAATAAACAG 1260

TGTTTTCCAT GAAAAAAAAA AAAAAAAAAA TTCCTGCGGC CGTCAAGG 1308

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1984 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CGTCAGAATC CACCTTGACG TGCTGTGCGT ATCTGTGAAC CTGGAGCRGT TACTTATTTT 60

GACAGATATC ACTTTGGGTC TTTTTACATT AAATTTCTTT TCTCTAAGGA ATATAAGACA 120

TACCCCATAG CTCTGYGTGA GCCAGCAATA CCGCTGCCCC CTGGCGACAG GGCAGACCAA 180

TGATGCCAGG CAGCTGTCAC ACGCTAGTAT TGGCTTCATT GTGATCTGAG CCCTGCACGC 240

TGGGCCTTCA GAATTAATGG CCAGCAGTGT CAGGGATGAG CCCGTCAGCC AGGGCACAGG 300

CCTGGYTCAC AGTCCTKCAC ACCTGCTGGC CTGGGGAGCT CCAGCCAGGC AGCGAGTCCT 360

GCCCCGCCCG CAGCTCCCTC CCACACCCCG CCTGGCCAAG ATGACTGCTT CAGGGGGCTT 420

TGGGGAAAGA ATTAGGAAGG GTCAGAACCA AACAATACCT GCTCATTTAC ACTGAGGATT 480

CAGGGCGGGA GACAGGAGCC TTGGGGTCCT GTTAAACCAC AGACAGTTAT GAACTGAAAG 540

TCATAACGGG GAGAGGTGCC TGGCTTCTAC CTGGGTGCTC AGGAATGTTC CTCGTCACCC 600

CTGCCACTCT GTGGTCGGTG CCCTGCTTCC TCCTCCACTC CTGGCCGCCT TCTCCAGCGC 660

CGCACACACA GATGCTCAGT CTCAGAGAGG CTGGCACGGC CTGGCAGTCT GAGAAAAGCG 720

TCAGTTAGGC ACACCTGCAG GCCCCTCGGT GGGACAGCGG CGGCCTTGGA GTTAGGAGCC 780

ACCCTGGGAG GTTGTGCCGG TGCCATGCTC CTCCCTGTGT CTTGTATGAA AGGGGCCACT 840

GTGTGTCTTC CTCCCCGGCG GGAGCCCCAC ATGTGTGCAC TGTAGGACAG CGGCCCCGAG 900

GTGGAAGCCT GGCTGGAGGG CTGCCCTATA GGTCTTCTCT TCCCGCCTCC CCTGCCATGC 960

AACCAGATGT GTTGTGAGTG GGCAGCGTGC CCCACGCTGG AGTAACTCCG CACGCTTCTG 1020

TCTTTCACGG TGGGCGCTCG GGGGGAGCCT GAGGAAAACC CCCTTAGGTA CCTGTGCGAG 1080

GCTGTGGAGT GCAGGCCAGA GCAGGGTGTG CGTANCCCCA GCACCCAGGT TCTTCTGTCA 1140

GACCCTGTGA CCTGCGAGCT GCTACTACTG TAAGGAGGGA AATGGATGAA TCTGGCTCGT 1200

TTTAAAATCA CGTTTTCTGA CGAATCCTTT GCCCCTTCAC CTTTACCCCG CCCGCACCCC 1260

TAGGCCCTCT CAGCCTTCCT ATCATCCCAC GTGTCTACCC AGACCCTTGT GCGGCCCATG 1320

CCCYGGGGGC GGCGTCCTGT CCCTGAGCTG GGAGGCGGCT TTGGATGGTC CGGGCGTCAA 1380

GAGCAGGGGT GGGCCGGGGA GGGGTCCTTT GCGGTGAGCT ATGTTTACAT GACACAGTGT 1440

GCCAAAGTGA CTTACTGCGG TTGCGTTAGT TTCTAGTCAT CAGGACTATC TCACCCTCCC 1500

ACTCCTGTTT TTAAAACTCA GAATTCTTTC CTAAGAGCCC TTCGAGCAAA GCGTGCCGAA 1560

GTTAGTTGTC TTCTCTGTGS TGGTCCTTTC TTATGTCCTC ATAAAAGCTC AGATGATGGT 1620

ATCTGTGAGT ATGTTTTGCA AATTCAAAAT ATAGTTTGGT AATTTTTTTT TCCAGTTGAT 1680

TTTTAAAAAG AACTGCTGTA CAGAGCTTGT ACTTTGTCCA TTTTATAGAT GGAAACCATC 1740

CTTGAAAATT GTTTAACTTA AATAAAGAGA AGATACTTTC TTCGTGCCGA ATTCGGCACG 1800

AGCCCAAACC CACTCCACCT TACTACCAGA CAACCTTAGC CAAACCATTT ACCCAAATAA 1860

AGTATAGGCG ATAGAAATTG AAACCTGGCG CAATAGATAT AGTACCGCAA GGGAAAGATG 1920

AAAAATTATA ACCAAGCATA ATATAGCAAG GACTAACCCC TATACTTTCT GCATAATGAA 1980

TTAA 1984

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2011 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGGCTGGGAT GTGAGGAGCG GCGGGTTCCG GGCTCCGGCT CTGGGTGGCG GCGGCTGTGA 60

GCGGCGGCAC TGCGGCGCAG CGGCGGAGGC CAGCGGGCGC CGTCGGNGCT GGCCCTGTCG 120

GCCGCGGGAT GAGGAAGCGG ACCGAGCCCG TCGCCTTGGA GCATGAGCGC TGCGCCGCCG 180

CGGGCTCGTC CTCCTCCGGC TCGGCCGCCG CGGCGCTGGA CGCCGACTGC CGCCTGAAGC 240

AGAACCTACG CCTGACGGGY CCGGCGGCGG CTGAGCCGCG CTGCGCACCG ACGCGGGAAT 300

GAAGCGGGCG CTGGGCAGGC GAAAGGGCGT GTGGTTGCGC CTGAGGAAGA TACTTTTCTG 360

TGTTTTGGGG TTGTACATTG CCATTCCATT TCTCATCAAA CTATGTCCTG GAATACAGGC 420

CAAACTGATT TTCTTGAATT TCGTAAGAGT TCCCTATTTC ATTGATTTGA AAAAACCACA 480

GGATCAAGGT TTGAATCACA CGTGTAACTA CTACCTGCAG CCAGAGGAAG ACGTGACCAT 540

TGGAGTCTGG CACACCGTCC CTGCAGTCTG GTGGAAGAAC GCCCAAGGCA AAGACCAGAT 600

GTGGTATGAG GATGCCTTGG CTTCCAGCCA CCCTATCATT CTGTACCTGC ATGGGAACGC 660

AGTACCAGAG GAGGCGACCA CCGCGTGGAG CTTTACAAGG TGCTGAGTTC CCTTGGTTAC 720

CATGTGGTCA CCTTTGACTA CAGAGGTTGG GGTGACTCAG TGGGAACGCC ATCTGAGCGG 780

GGCATGACCT ATGACGCACT CCACGTTTTT GACTGGATCA AAGCAAGAAG TGGTGACAAC 840

CCCGTGTACA TCTGGGGCCA CTCTCTGGGC ACTGGCGTGG CGACAAATCT GGTGCGGCGC 900

CTCTGTGAGC GAGAGACGCC TCCAGATGCC CTTATATTGG AATCTCCATT CACTAATATC 960

CGYGAAGAAG CTAAGAGCCA TCCATTTTCA GTGATATATC GATACTTCCC TGGGTTTGAC 1020

TGGTTCTTCC TTGATCCTAT TACAAGTAGT GGAATTAAAT TTGCAAATGA TGAAAACGTG 1080

AAGCACATCT CCTGTCCCCT GCTCATCCTG CACGCTGAGG ACGACCCGGT GGTGCCCTTC 1140

CAGCTTGGCA GAAAGCTCTA TAGCATCGCC GCACCAGCTC GAAGCTTCCG AGATTTCAAA 1200

GTTCAGTTTG TGCCCTTTCA TTCAGACCTT GGCTACAGGC ACAAATACAT TTACAAGAGC 1260

CCTGAGCTGC CACGGATACT GAGGGAATTC CTGGGGAAGT CGGAGCCTGA GCACCAGCAC 1320

TGAGCCTGGC CGTGGGAAGG AAGCATGAAG ACCTCTGCCC TCCTCCCGTT TTCCTCCAGT 1380

CAGCAGCCCG GTATCCTGAA GCCCCRGGGG GCCGGCACCT GCAATGCTCA GGAGCCCAGY 1440

TYGCACCTGG AGAGCACCTC AGATCCCAGG TGGGGAGGCC CCTGCAGGCC TGCAGTGCCC 1500

GGAGGCCTGA GCATGGCTGT GTGGAAAGCG TGGGTGGCAG GCATGTGGCT CTCCTTGCCG 1560

CCCCTCAACC TGAGATCTTG TTGGGAGACT TAATGGCAGC AGGCAGCCAT CACTGCCTGG 1620

TTGATGCTGC ACTGAGCTGG ACAGGGGGAG TCCGGGCAGG GGACTCTTGG GGCTCGGGAC 1680

CATGCTGAGC TTTTTGGCAC CACCCACAGA GAACGTGGGG TCCAGGTTCT TTCTGCACCT 1740

TCCCAGCACA TGCAGAATGA CTCCAGTGGT TCCATCGTCC CCTCCTGCCC TGTGTACCTG 1800

CTTGCCTTTC TCAGCTGCCC CACCTCCCCT GGGCTGGCCC ACTCACCCAC AGTGGAAGTG 1860

CCCGGGATCT GCACTTCCTC CCCTTTCACC TACCTGTACA CCTAACCTGG CCTTAGACTG 1920

AGCTTTATTT AAGAATAAAA TCGTGGTGGT GGTCAAAAAA AAAAAAAAAA GGGGGCCGCT 1980

TAAGGGTCCA NNTTAAGNAA GGGGAATTGG A 2011

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1380 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GGACTGCGCG GCCGTTGGGG TGCAGCGGCG CCAGTCGGCG ACGAGGGGCC CCCGGGAGTT 60

GCTGGACTGA GACATGAGCC TCCAACTGTG TGGTTGGGCT CGGTAGCACA TCGTGGGACT 120

TGGGTGTGCG CCCACAGATG GTTTGGCCCT GCAGTGACCA GAGCAGCCCA AGCCGCCACC 180

ATGGTGAAAT TGCTAGTGGC CAAAATCCTG TGCATGGTGG GCGTGTTCTT CTTCATGCTG 240

CTCGGCTCCC TGCTCCCCGT GAAGATCATC GAGACAGATT TTGAGAAGGC CCATCGCTCG 300

AAAAAGATCC TCTCTCTCTG CAACACCTTT GGAGGAGGGG TGTTTCTGGC CACGTGCTTC 360

AACGCTCTGC TGCCCGCTGT GAGGGAAAAG CTCCAGAAGG TCCTGAGCCT CGGCCACATC 420

AGCACCGACT ACCCGCTGGC CGAAACCATC CTCCTGCTGG GCTTCTTCAT GACCGTCTTC 480

CTGGAGCAGC TGATCCTGAC CTTCCGCAAG GAGAAGCCGT CCTTCATCGA CCTGGAGACC 540

TTCAACGCCG GATCGGACGT GGGCAGCGAC TCGGAGTATG AGAGCCCCTT CATGGGGGGC 600

GCGCGGGGCC ACGCGCTGTA CGTGGAGCCC CACGGCCACG GCCCCAGCCT GAGCGTGCAG 660

GGCCTCTCGC GCGCCAGCCC CGTGCGCCTG CTCAGCCTGG CCTTCGCGCT GTCGGCCCAC 720

TCGGTCTTTG AGGGCCTGGC CCTGGGCCTG CAGGAGGAGG GGGAGAAAGT GGTGAGCCTG 780

TTCGTGGGGG TGGCCGTCCA CGAGACACTG GTGGCCGTGG CCCTGGGCAT CAGCATGGCC 840

CGGAGTGCCA TGCCCCTGCG GGACGCGGCC AAGCTGGCGG TCACCGTAAG CGCCATGATC 900

CCCCTGGGCA TCGGCCTGGG CCTGGGCATT GAGAGCGCCC AGGGCGTGCC GGGCAGCGTG 960

GCGTCCGTGC TGCTGCAGGG CCTGGCGGGC GGCACCTTCC TCTTCATCAC CTTCCTGGAR 1020

ATCCTGGCCA AGGAGCTGGA GGAGAAGAGT GACCGTCTGC TCAAGGTCCT CTTCCTGGTG 1080

CTGGGCTAMA CCGTCCTGGC CGGGATGGTC TTCCTCAAGT GGTGAGCGGC CTCGCCATTG 1140

TCCCTGCCGC CGGAGCCCGC GGGAGCCCCG GGCCGGACAC AGGCGNGTCC CCCGGCCGCG 1200

CGTCCCCCAA GAGCGAGCAC TTGGCCCTGG GCCACCACCT GTGCACAAGG GGCCTCCCGG 1260

GACCAGGCTG TGCCCCCGAT CCTACACCCT GAGCCTCAGA GCACTGCTAC TTTTTAAAAT 1320

ACTTCTTTCT CTTAAAAGTC TTTACCAAAA AAAAAAAAAA AAAAWAAGGG GGGCCCGGTA 1380

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2041 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGATAATGAA ACTTTTTATA TCTGAATTGC ACAAATCCCC ACCAAGTAAT TTCAGCATAA 60

AGCAAACCAA ATGGCAGACG AAATGATGTT AGTACCTTGT AGTTTACTTC ATAATACATA 120

ACCTCTAATC CATGTCCATC TTCTTGTTTC TGCTTTTAAT TCTGCCATTT CCTCTTAAAA 180

TTTAGGGTTT AACAAAGACA ACAGGGCTGT TTGCCAAATC GCAGCTATTA ATTCACAGCA 240

TAGAAAAGTC AAAGCTATAG CAAAAAATTG CTAATCTGCA CAACTTTAAA AAATAGTTCA 300

GTACATTTTT GTTATAAAAT TCATTTACAG GAGGTTATTC ACATGTACTT GTCAAATTTA 360

CTCCTGATAA TTCACAAAAA CATACAACTC AACAAACTGT GCACAATAAA TCCAAGGCAA 420

ATTATATACA AAGAAACAAA ACAAGCTTTT AAGTAGCACA TATTCATTTG AAATAACTAA 480

TATTGAAAGA AGACAGGGAA CTTTCTTTTA ATGCCATGGC AAAGACGAAG CGAAGAGCCA 540

CACTTCACAC CTTGTAAAAA GAATAGCCCT GTTCAACAAC NCTGCGCTGA CAGCCACATC 600

AGGAGGGGCC ACGGTGAACA TAGGAAATGG CTTTGGCAAA TACTTGTACC AACTGGAACG 660

AGTGAAGTTT CAAAAGTAAT GTGAGGTACA ACTGCATTCC GCTGTGAAAG GCCGTCACAG 720

GACACAGGCT CGTCTGTTAG AAAGGATGAT CTAGTTCTAC CATTAATTCT TGCAGAATCA 780

GATCTGCTGA GTGGTGAACC AACAGGTGAA CACAACGTAA GAACAGGCAT CAAACTTCAC 840

TGGAAATAAT ATTGTTCAGT GTGTGGCGGC AAAATATGCA TTTTAGAGAA AACTTATTTC 900

TCAAATCATG TGTTAATAGT ATTAACATGA GCAGCGTGAG AGACATCCTG ACCCCAACGT 960

TTTTGCCATG CCTCCCTTTA GAAGCTTAGG AGTTTGTACA TTCCCTAAGT GGTCAGCACT 1020

ACAAGTGTCT GCTAAAATGG GCACTTCATC AAGATAACAG GAAAGCAAAC TTAAAGTAAC 1080

GAGATTTCTT CCCAAAGGCA CATGGAAGAA GCTGATAGAG CCCTTGACCC AGACAGAATG 1140

GGACCCATCC CTACCCGTCC TGAACTGTCG CACACTGCAT GGCCAAAGAC AAACTCTCCA 1200

CCCCCACAGA GGAAGCAGTG GCTGACTCTG GGGACAAAGC ACTCCAGGAA GTCACCTGCT 1260

CCCTGGGTTT CAGGAGTATT CAGTTGACCG TCTGGACACC AGTGAGGGAG ACACAGGTAA 1320

TGAAATCAAA TGCTCAAACT TTTGGCAACC GAAAGTTGTT TTTTAAAGCT CTTTATAATC 1380

TGCTCAAGTA GAATTTCTAA CACAAAACCC TTTTTTGCTT AAAAAGCAGA TGACAAAGGA 1440

AATGTCAAAT AATGCACATG AATCTTCAGC TATTTTCCTA CCCCCAAATG AGATATGGGG 1500

CTGCACAGCA TTCACTACAG ATCCCTAGTT TTTACAACTG TCAACTGTAC ATTCTCATGT 1560

TTAGGATACT CCAGGTTNCT GCGTGGATAT TTGGAAAACT GGACAAAATC AGGCAGCCCA 1620

CCCCTCCCTT GTCCCGAGCT TCCTCGATCT CCATTCACCA TGACCAATTT TTTCCCCCAC 1680

AAAAGCACTA TCACCTCTAA TAGTAGCTGG AAACACCTAT CAGATATTCT AAACAGCATG 1740

TATTTTTACC AGGTAGATGA TTTCTGAAGA TCACAGGAAG TCTACCACTC TCTTCCCAGT 1800

TCTGAACTCC TCTGGTTACG CTTCATTTTA AATCGGGTGC TTCTTTCCCG CCCAAAATAT 1860

TCTATTTGGC CTGCCCCAGT GGTTCCCAAG CTGCCTGGTG ATCAGAATCC CTTGGGAAAT 1920

GTTGAACACA CAGCTTCCCA GGCTTTCTGG AAAACTASTT AGATCTGTCT GACAATCTGT 1980

AAGCTGAGGC GATTCTTCCA CAGCTGACCC AGCGCTAATC TASSATTTGG CAACCAAATG 2040

A 2041

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1875 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCTAAATAAA AGGGTCTAAA ACTCAGCTTC TGAGTTTTTA AAATCACGGT CTCCAGGTAC 60

CAATAAATGC TACAGTTTGC CTTATGATGT TAACATAAAA CACTTAGTAG AAGGACAATA 120

TTTCCATGAA AATAATGTTT TTCAATATTA AGAAGTTACT ACTCAAATTT TCACAGTAAG 180

CCATTTAGGG TATGTTTGGC TATTTTTATA AGGACATGAG AGATTATGTC ATAATTTTGT 240

TGTGGAAGTC TCACTCTTGG CTAACTTAAA AGCATTGTGG ATAGTAGCAG TTACTAGTTC 300

CAGGTTGTCA TATTTACAGG AAAATATGTA TATGGTGAAA GGCCACCGTG TTTAATTACT 360

ATAATGATGT AGAAAAGATT CCCGTGTGAA TTTTTTTTTT GAAAGTCTAA AAAATGTATG 420

CTGTAAAAAT TTGCTGCAGT GTAATTTTGC ATTCTCTTTA AACTGATTGA GGTCACAGTA 480

TTTTATTATT TGGGGTCCTC ACCACAGGAA ACACTGCGAT ACAGGGGCAA AAGAGATGGC 540

AGTGCCAATT TAAATTAATA CAACAAAATC AATGCAGCAC CAACCAAGAC TGCCAGGTCT 600

GGTGTCATGG GTATGCCCAG AGCCCAGGAG TTCAGAAGGG CCCTAAGCCT GATTTAATGC 660

TCTGCTGTTG ATGTCTTGAA ATTCTTAACA ATTTTTGAAC AAGGGGCCTG CGTTTTCACT 720

TCGCACTGGG CCTTGCAAAT TACATAGCGA GTGCTCATAA AAGAACTCAG AAACGTGGTA 780

CCTCTCTTCC TGGTGGATAC AAATAAAGAA ATCTGGATCC AAAGTTGAAA GTTGCTGGCG 840

ATATCATTCA AGTAGGACTC TAAATAGTGG ATTAAGATGA GGGTGGGCCT GGGTGAAGAT 900

TCTTTCCAGC TTTAAAAGAA AGTGACTTCA AAAACTGACT GCAAATATTG ACGATGGTTT 960

CTGCTGGAGG AAAAGAAACA GCTTGAATAC AGACAGGCTT TTTTATTACG GTACTGATAT 1020

ATTGACCTTA AACTTGCTGA GGAACTGAAC TAACGTCCTC CAGTGACCGT GGAATTCCAT 1080

CTCAGCTCCA GGAACATGCA GATACCTGCA AAGAGACACG CATATATGCT GGCATACATG 1140

TGCATTTGGT GTTGGGAAGT TGACCATCTG GTCTATCTTA ATAAAATGGT AAAAAGCACA 1200

CCAAGACAAT GATGGGGGCA GGAGGATGTT TTTGAAAACA GCGCTTCTCA ACCAGTGCTC 1260

GATTTTGCCC CCCAGGAGAC ATTTGGCAAT GCAATGGCAA CTTTTGGTTG TCGCAGCCGG 1320

GGAAGGGAAG CTACCAGCAT CTAGTGCGTA GAGGTCATGG ACGCCGTTAA ACATCCTACA 1380

GTGCAAGCGC ANCNCCCGAC CACGAAGAGT TGTCTTGCTC AAATATCAAC AGTGCTGCAG 1440

TGTAGAAACT TGATCGTTGG TTTTCTTTTA ATGCAAAACT CTCATAAAAA CCTTTCACTT 1500

TTCCTGTCAT TGATTATATG CTTGATACAC CCAAAAAGAA AAGGGGAGGG GCACCAATTC 1560

ACCTACACTC CAGTGGCTCC ATCACCTTTA AAAATATTTA TAAAATAGTT CCAAAAATCT 1620

GATATCTGAA AAGCAATCCA AGCCTGTGTA AATGGGAATC ACTGATAAGT ATCATCATCT 1680

GTATCAGCTT GGCTTGGACA TGAAAAATTG ATTCTCTTTA TGTCACTCCT TGCACCTGGA 1740

CAAATTCAAT CCCCGGTACT TAAGTCACAC TGCCAAGCCC TCGGCCCTGA CTATTGTCTT 1800

GATTGCTGTT CCTTTCTGGT TCAAAATAAA ATCATTTTTG TGGCACCAAG AAAAAAAAAA 1860

AAAAAAAAAA CTCGA 1875

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2432 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

TTAGATGCTG TTTTAGAATA CCTCCCAAAT CCATCTGAAG TCCAGAACTA TGCTATTCTC 60

AATAAAGAGG ATGACTCAAA AGAGAAAACC AAAATCCTAA TGAACTCCAG TAGAGACAAT 120

TCCCACCCAT TTGTAGGCCT GGCTTTTAAA CTGGAGGTAG GTCGATTTGG ACAATTAACT 180

TATGTTCGCA GTTATCAGGG AGAGCTAAAG AAGGGTGACA CCATCTATAA CACAAGGACA 240

AGAAAGAAAG TACGGTTGCA ACGGCTGGCT CGCATGCATG CCGACATGAT GGAGGATGTT 300

GAGGAAGTAT ATGCCGGAGA CATCTGTGCA TTGTTTGGCA TTGACTGTGC TAGTGGAGAC 360

ACATTCACAG ACAAAGCCAA CAGCGGCCTT TCTATGGAGT CAATTCATGT TCCTGATCCT 420

GTCATTTCAA TAGCAATGAA GCCTTCTAAC AAGAACGATC TGGAAAAATT TTCAAAAGGT 480

ATTGGCAGGT TTACAAGAGA AGATCCCACA TTTAAAGTAT ACTTTGACAC TGAGAACAAA 540

GAGACAGTTA TATCTGGAAT GGGAGAATTA CACCTGGAAA TCTATGCTCA GAGGCTGGAA 600

AGAGAGTATG GCTGTCCTTG TATCACAGGA AAGCCAAAAG TTGCCTTTCG AGAGACCATT 660

ACTGCCCCTG TCCCGTTTGA CTTTACACAT AAAAAACAAT CAGGTGGTGC AGGCCAGTAT 720

GGAAAAGTAA TAGGTGTCCT GGAGCCTCTG GACCCAGAGG ACTACACTAA ATTGGAATTT 780

TCAGATGAAA CATTCGGATC AAATATTCCA AAGCAGTTTG TGCCTGCTGT AGAAAAGGGG 840

TTTTTAGATG CCTGCGAGAA GGGCCCTCTT TCTGGTCACA AGCTCTCTGG GCTCCGGTTT 900

GTCCTGCAAG ATGGAGCACA CCACATGGTT GATTCTAATG AAATCTCTTT CATCCGAGCA 960

GGAGAAGGTG CTCTTAAACA AGCCTTGGCA AATGCAACAT TATGTATTCT TGAACCTATT 1020

ATGGCTGTGG AAGTTGTAGC TCCAAATGAA TTTCAGGGAC AAGTAATTGC AGGAATTAAC 1080

CGACGCCATG GGGTAATCAC TGGGCAAGAT GGAGTTGAGG ACTATTTTAC ACTGTATGCA 1140

GATGTCCCTC TAAATGATAT GTTTGGTTAT TCCACTGAAC TTAGGTCATG CACAGAGGGA 1200

AAGGGAGAAT ACACAATGGA GTATAGCAGG TATCAGCCAT GTTTACCATC CACACAAGAA 1260

GACGTCATTA ATAAGTATTT GGAAGCTACA GGTCAACTTC CTGTTAAAAA AGGAAAAGCC 1320

AAGAACTAAC TTTGCTTACT GTGAGTTGAC TGACTCTAAT TGAATCTGCG TGGTTTTGAT 1380

ACTTTGATGG ATTCCAGTGG AATAAATTCA GGCTGCTGAA ACAAGAAATT CTGAGCCCAG 1440

GAAGCGGGCT CTTCTTTCTT CAAAAGAAGC CCTTCTTGTT CATATTCAGG AGCTTCTGTT 1500

ATATTCAAAG GTAATTCTAT GTCTATCTCA ACTCTATTGA TTGGTTTTAT AGTTCATTGA 1560

AAATCCTCAA ATAAAATATA ATTATTACTG AAATATGTTT AATATTTAAG GGGAAAAGAG 1620

ACTAATTTCA GTTATACTTT TAAGCTTAGA ATGTATGTTC ATTTCCAAAT TTTGTATCAT 1680

AAGAGTTTTC AACATAGAGA AAAGCTGAAA AAATGCAAAG AATAACCACA TACTTTCCAT 1740

CTACCTTCCT TTGTTAACGG GTTGTTTATC ATATAATAAT TTGTTTTGTC ATATTTGCTT 1800

TCACTGTCTA TTATCTGTTT AAGTCTCATA ACTCTATTTT TAGTTTGCTG AAGACTTGAA 1860

AGTGAATCGC ATATATCATG ACACTTCTTG GAGTGTCATT AATGGGCAGG CTTTTCTGTT 1920

GAAGAGTGGA TTCCGTATGT TCTTCATAGA GAGTGTTTTT CAGATTCTTC ATTGGGATAT 1980

TAAAATATTA GCCAAATTTC YCTCTGTTTT ATATATGYCA GTTTATTTCA GTTTGTGGTT 2040

TCTGCAAATT TGTAACTGCC TCTGTTTTAG GAGTATAAGT ATTACTTCCT TGTGGTCTAT 2100

TGTGAAGTAA AAAGTAGACC CTTGCATATA CTATTCTTGT TTGTGTTCAT CTTAATGTTT 2160

TTGTACAGCT AAATCAAATG TAATTTATAG AGTTAGTTTC ATCAACCTAA TGAATGCTAG 2220

TTAAATTTGA ATTCCTTGGA ATTTATCGTA TATTGTATTC ACTGAGATTA TGAAGGGACA 2280

AATGTTAATC TTTTGTTTCC AGAAAAAGTT GGGCTTTCCC AAGCAGTTCT ATTACCCGGT 2340

TCAGAATTGC TTCATCCAAA AATCATCTGA TGGTATAGAT GGATCCTAGT CCTTTTCATT 2400

ACCTGATGGT AGAAATAAAA TAATTGATTT TA 2432

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1269 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TCGACCCACG CGTCCGGGCG GCGGCAGCAT GGCGGCGGGG GCGGCTGAGG CAGCTGTAGC 60

GGCCGTGGAG GAGGTCGGCT CAGCCGGGCA GTTTGAGGAG CTGCTGCGCC TCAAAGCCAA 120

GTCCCTCCTT GTGGTCCATT TCTGGGCACC ATGGGCTCCA CAGTGTGCAC AGATGAACGA 180

RTTATGGCAG AGTTAGCTAA AGAACTCCCT CAAGTTTCAT TTGTGAAGTT GGAAGCTGAA 240

GGTGTTCCTG AAGTATCTGA AAAATATGAA ATTAGCTCTG TTCCCACTTT TCTGTTTTTC 300

AAGAATTCTC AGAAAATCGA CCGATTAGAT GGTGCACATG CCCCAGAGTT GACCAAAAAA 360

GTTCAGCGAC ATGCATCTAG TGGCTCCTTC CTACCCAGCG CTAATGAACA TCTTAAAGAA 420

GATCTCAACC TTCGCTTGAA GAAATTGACT CATGCTGCCC CCTGCATGCT GTTTATGAAA 480

GGAACTCCTC AAGAACCACG CTGTGGTTTC AGCAAGCAGA TGGTGGAAAT TCTTCACAAA 540

CATAATATTC AGTTTAGCAG TTTTGATATC TTCTCAGATG AAGAGGTTCG ACAGGGACTC 600

AAAGCCTATT CCAGTTGGCC TACCTATCCT CAGCTCTATG TTTCTGGAGA GCTCATAGGA 660

GGACTTGATA TAATTAAGGA GCTAGAAGCA TCTGAAGAAC TAGATACAAT TTGTCCCAAA 720

GCTCCCAAAT TAGAGGAAAG GCTCAAAGTG CTGACAAATA AAGCTTCTGT GATGCTCTTT 780

ATGAAAGGAA ACAAACAGGA AGCAAAATGT GGATTCAGCA AACAAATTCT GGAAATACTA 840

AATAGTACTG GTGTTGAATA TGAAACATTC GATATATTGG AGGATGAAGA AGTTCGGCAA 900

GGATTAAAAG CTTACTCAAA TTGGCCAACA TACCCTCAGC TGTATGTGAA AGGGGAGCTG 960

GTGGGAGGAT TGGATATTGT GAAGGAACTG AAAGAAAATG GTGAATTGCT GCCTATACTG 1020

AGAGGAGAAA ATTAATAAAT CTTAAACTTG GTGCCCAACT ATTGTAAGAA ATATTTAATT 1080

ACATTGGGAG CAGTTCATGA TTTAGTCCTC AGAAATGGAC TAGGAATAGA AAATTCCTGC 1140

TTTCTCAGTT ACATGTTTTG TGTATTTCAC AATGTCGTGC TAAATAAATG TATGTTACAT 1200

TTTTTTCCCA CCAAAAATAG AATGCAATAA ACATCTTCAA ATTATTAACA ATAAAAAAAA 1260

AANAAAAAA 1269

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 762 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGCACGAGCG AAGATCAAAG TGTTGACAGG CATGGGTTCT TCTGAGACAT CCCTCCTTGG 60

CTTGCAATTG GTCACCTTCT TGCTTCTTCA CATGGTCCTC CTTCTGTGTG TGACTGTGTC 120

CAAATTTCCT TTCTGTAAGG ACACAGCCAT ATTGGATTAG GCCCACCCTA TTGACCTCAT 180

CTTAACTTAT TTACTCCTTT AAAAACCCTG ACTCCTTATA CAGTCACACT CCGAGGTACT 240

GGGGGATTAG GATTTCAATG TATGAATTTT GGGAGGTGAG AAGGACANAA TTTCAGCCAA 300

TACCAGTTAA ATGGATTTAG TAATTCAAAC ACAGGGGATT GGAATACGGC AGATTTTTAA 360

GGGNNTGGGA ATTGAAGCCA GAATTTNGGA AGGGNTTTAG AACTGATGGG AGGGCAGGTG 420

NCTGGGTCCN GGGNGATTTT GGAAAAAGAT TTTTCAGGCC AGGTGAGGTG GCTGATTCCT 480

GTAATCCCAG CACTTTTGGA GRCCGAGGCT GGCAGATCAC TTGTAGGCCA GGAGTTTGAG 540

ACCAGTCTGG GAAACATGGC AAAACCCTGT CTCTACAAAA ATTACAAAAT ATCAGCCAGA 600

AGTGGTGGCT TGTGCCTGTA GGCCCAGCTC CTCTGGAGGC TGAGGTGGGA GGATCACTGG 660

AGCCTGGGAA GTCAAGTCTG CAGTGAGCAA AGATCTGTGC CTCTGCACCC CAAGCTGGAC 720

AACAGAGCAA GACCCTGTCT CCAGAAAAAA AAAAAAAAAA AA 762

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2888 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

TCGACCCACG CGTCCGGGAT GAGGCCCGGC CTCTCATTTC TCCTAGCCCT TCTGTTCTTC 60

CTTGGCCAAG CTGCAGGGGA TTTGGGGGAT GTGGGACCTC CAATTCCCAG CCCCGGCTTC 120

AGCTCTTTCC CAGGTGTTGA CTCCAGCTCC AGCTTCAGCT CCAGCTCCAG GTCGGGCTCC 180

AGCTCCAGCC GCAGCTTAGG CAGCGGAGGT TCTGTGTCCC AGTTGTTTTC CAATTTCACC 240

GGCTCCGTGG ATGACCGTGG GACCTGCCAG TGCTCTGTTT CCCTGCCAGA CACCAMCTTT 300

CCCGTGGACA GAGTGGAACG YTTGGGAATT CACAGCTCAT GTTCTTTCTC AGAAGTTTGA 360

GAAAGAACTT TCYAAAGTGA GGGAATATGT CCAATTAATT AGTGTGTATG AAAAGAAACT 420

GTTAAACCTA ACTGTCCGAA TTGACATCAT GGAGAAGGAT ACCATTTCTT ACAMTGAACT 480

GGACTTCGAG CTGATCAAGG TAGAAGTGAA GGAGATGGAA AAACTGGTCA TACAGCTGAA 540

GGAGCCTTTT GGTGGAAGCT CAGAAATTGT TGGACCAGCT GGAGGTGGAG ATAAGAAATA 600

TGACTCTCTT GGTAGAGAAG CTTGAGACAC TAGACAAAAA CMATGTCCTK GCCATTCGCC 660

GAGAAAYCGT GGCTCTGAAG ACCAAGCTGA AAGAGTGTGA GGCCTCTAAA GATCAAAACA 720

CCCCTGTCGT CCACCCTCCT CCCACTCCAG GGAGCTGTGG TCATGGTGGT GTGGTGWACA 780

TCAGCAAACC GTCTGTGGTT CAGCTCAACT GGAGAGGGTT TTCTTATCTA TATGGTGCTT 840

GGGGTAGGGA TTACTCTCCC CAGCATCCAA ACAAAGGACT GTATTGGGTG GCGCCATTGA 900

ATACAGATGG GAGACTGTTG GAGTATTATA GACTGTACAA CACACTGGAT GATTTGCTAT 960

TGTATATAAA TGCTCGAGAG TTGCGGATCA CCTATGGCCA AGGTAGTGGT ACAGCAGTTT 1020

ACAACAACAA CATGTACGTC AACATGTACA ACACCGGGAA TATTGCCAGA GTTAACCTGA 1080

CCACCAACAC GATTGCTGTG ACTCAAACTC TCCCTAATGC TGCCTATAAT AACCGCTTTT 1140

MATATGCTAA TGTTGCTTGG CAAGATATTG ACTTTSCTGT GGATGAGAAT GGATTGTGGG 1200

TTATTTATTC AACTGAAGCC AGCACTGGTA ACATGGTGAT TAGTAAACTC AATGACACCA 1260

CACTTCAGGT GCTAAACACT TGGTATACCA RGCAGTATAA ACCATCTGCT TCTAACGCCT 1320

TCATGGTATG TGGGGTTCTG TATGCCACCC GTACTATGAA CACCAGAACA GAAGAGATTT 1380

TTTACTATTA TGACACAAAC ACAGGGAAAG AGGGCAAACT AGACATTGTA ATGCATAAGA 1440

TGCAGGAAAA AGTGCAGAGC ATTAACTATA ACCCTTTTGA CCAGAAACTT TATGTCTATA 1500

ACGATGGTTA CCTTCTGAAT TATGATCTTT CTGTCTTGCA GAAGCCCCAG TAAGCTGTTT 1560

AGGAGTTAGG GTGAAAGAGA AAATGTTTGT TGAAAAAATA GTCTTCTCCA CTTACTTAGA 1620

TATCTGCAGG GGTGTCTAAA AGTGTGTTCA TTTTGCAGCA ATGTTTAGGT GCATAGTTCT 1680

ACCACACTAG AGATCTAGGA CATTTGTCTT GATTTGGTGA GTTCTCTTGG GAATCATCTG 1740

CCTCTTCAGG CGCATTTTGC AATAAAGTCT GTCTAGGGTG GGATTGTCAG AGGTCTAGGG 1800

GCACTGTGGG CNTAGTGAAG CCTACTGTGA GGAGGCTTCA CTAGAAGCCT TAAATTAGGA 1860

ATTAAGGAAC TTAAAACTCA GTATGGCGTC TAGGGATTCT TTGTACAGGA AATATTGCCC 1920

AATGACTAGT CCTCATCCAT GTAGCACCAC TAATTCTTCC ATGCCTGGAA GAAACCTGGG 1980

GACTTAGTTA GGTAGATTAA TATCTGGAGC TCCTCGAGGG ACCAAATCTC CAACTTTTTT 2040

TTCCCCTCAC TAGCACCTGG AATGATGCTT TGTATGTGGC AGATAAGTAA ATTTGGCATG 2100

CTTATATATT CTACATCTGT AAAGTGCTGA GTTTTATGGA GAGAGGCCTT TTTATGCATT 2160

AAATTGTACA TGGCAAATAA ATCCCAGAAG GATCTGTAGA TGAGGCACCT GCTTTTTCTT 2220

TWCTCTCATT GTCCACCTTA CTAAAAGTCA GTAGAATCTT CTACCTCATA ACTTCCTTCC 2280

AAAGGCAGCT CAGAAGATTA GAACCAGACT TACTAACCAA TTCCACCCCC CACCAACCCC 2340

CTTCTACTGC CTACTTTAAA AAAATTAATA GTTTTCTATG GAACTGATCT AAGATTAGAA 2400

AAATTAATTT TYTTTAATTT CATTATGRAC TTTTATTTAC ATGACTCTAA GACTATAAGA 2460

AAATCTGATG GCAGTGACAA AGTGCTAGCA TTTATTGTTA TCTAATAAAG ACCTTGGAGC 2520

ATATGTGCAA CTTATGAGTG TATCAGTTGT TGCATGTAAT TTTTGCCTTT GTTTAAGCCT 2580

GGAACTTGTA AGAAAATGAA AATTTAATTT TTTTTTCTAG GACGAGCTAT AGAAAAGCTA 2640

TTGAGAGTAT CTAGTTAATC AGTGCAGTAG TTGGAAACCT TGCTGGTGTA TGTGATGTGC 2700

TTCTGTGCTT TTGAATGACT TTATCATCTA GTCTTTGTCT ATTTTTCCTT TGATGTTCAA 2760

GTCCTAGTCT ATAGGATTGG CAGTTTAAAT GCTTTACTCC CCCTTTTAAA ATAAATGATT 2820

AAAATGTGCT TTGAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 2880

GGGCGGCC 2888

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ACGAGTGCGG GCAGCAGCAG CCCCGGCACG MGGGAGAGAG ACAAAGCATG GAGGACACAA 60

CAATGGGAGG AAAGGCGGAC TCTCAGGAAC TTCATTCTTC ACGTGGTTTA TGGTGATTGC 120

ATTGCTGGGC GTCTGGACAT CTGTACCTGT CGTTTGGTTT GATCTTGTTG TTGATGAGCA 180

GATTACTAGC CAAAGCAAAG GACTTCCGTT ATAACTTATC AGAGGTGCTT CAAGGAAAAC 240

TAGGAATCTA TGATGCTGAT GGTGATGGAG ATTTTGATGT GGATGATGCC AAAGTTTTAT 300

TAGGCCTGAC CAAAGATGGC AGTAATGAAA ATATTGATTC TCTTGAGGAA GTCCTTAATA 360

TTTTAGCAGA GGAAAGTTCA GATTGGTTTT ATGGTTTCCT CTCATTTCTC TATGATATAA 420

TGACTCCTTT TGAAATGCTA GAAGAAGAAG AAGAAGAAAG CGAAACCGCA GATGGTGTTG 480

ATGGTACGTC ACAGAATGAA GGGGTTCAGG GAAAGACTTG TGTCATATTG GATTTACATA 540

ACCAGTAACC TTGATTCAGG GACTGAAGTC ATTGGCTAAT GAACACCTGA AGCAGCCTCC 600

TTTTTCTTTT CTTTCCTTGG CTTATGCAGG GCTTAATGTG CAGTGGGGTG GTTGTGATCT 660

TACCGTGCAA GTCAACCATG TGATCTTGCC CAGTACAGCT ACTAGCCTAG TCCCTTGCTC 720

GCTCAGCTCC CCCAACTTCT ATTGAAGAAA ATGGTACTCC TCATTCTTGT AGTCAGCTAC 780

AAAGTACACT GAAAATGATG TTCTTGGTGG TATAATTGGT TTCTGTATCG TTTTGTTTCA 840

ACTCATGTAT TCACTGAACT AAATTTGGAC ACTTAACAGC AAATTGTGTT GTGGTTAACC 900

CTTGATGCTT GTCTTTCTAA CACACTATTA ATTATGATGA TTCTAATGGA TTTCATTATA 960

AAAATATTTC TGGCATGATT TTTAAGTTAA ATGCTTCTCT GTTCTTTAAC ATGACTGATG 1020

TATAAAATGA TGGTTCTTTT ACTAAGCTGA TATTTTTTAT TGTAATTTGT TTAGGTTTGT 1080

CAGATAGGTT CATACAAAAT TAAAAGTAAA ATTCTGTGTT AATGGTGCTT TTAAAATAAT 1140

TTAAAAATAA CTCCATGTTT TTGCCTTAGA GTAAGTTAAC TTACTGTTTT CAGATAGTAG 1200

CATGACATAT TTCTGTCTGT GAAAGCAAAA TTTATTTTAA ATTTTATTTC CAAATATACA 1260

TCCAGAGAAA GTAATTTGTA TTTTTTTTAA AGTAGGCATA TTACACAAGA GGGAACATGT 1320

GAATATGTAT CTTAATGTTG TACATAGGGA AATTATTCAT CCTAAAAAAA AAAAAAAAAA 1380

AA 1382

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1656 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

CTTTATTGTT TTATATAGAT GGGATTATAC TAATTGTTAT ATCCTAACAA TTAATAGTTA 60

TATACTGACT GTATAAATGT TATACTCACA TTTATATAGA TGGGAATATA CTATTCCTTT 120

TTTGTTGTTA CTTATCATGG CCTCCTCTCC CAGCCTGTTC TGTCTGCCTC GGTCTCTGAA 180

GTCCGTGTAG GATTCACAGT ATCATGGGGG ACAGAAGTGC TATAGGTTGT TGAACCCTCC 240

TGTCTTGACG CATATTTCTA GTCATTTCCA GATGCCTACA TAATCCAGTA AAACTCCTTC 300

TGTGTAGATC TTCTGATGTA CTTGTATATG CAGATTTTTA GCAAATATTC CTAAAATTGA 360

ATTCCTAGAA TTGCTGGGCT GGTAGGGTTT TTAATTCTGA TATCGTGAAA TCGATCGCTA 420

GAGAAATTTT GGTACCTCTC TTTGGATCAA AGGCATCAGC ATTTTTAAAT GAAGCTTGAA 480

CTGATTTGTG TACTGGAATC CATATCAAAC TACACAAATT TGCTAAATCC CTAACGAAAA 540

CAGTAATGTT TCANCAAACT GTGAGCAGAC CCAAAGGGCG TTGATGGTAT TAATTATACA 600

TCAGCCTGAG TGGAAGTCAA ACCAAGCTAG TTTTAGAAGC TATCCACAAC TGGTAAAGGT 660

AAACCTGAAT CTTTTTAAAA ATTGTGATAA AGTACACGTA ACATAAAATT CACCATTTTA 720

ACCATTTTTA AGTATACAGT TCAGTGACAT TTAAGTCCAT CCACACTGCT GTGAAACTGA 780

AAGCTGGATC TTAATTTCTA GTCTCTAAAC TGAACTGAAA TCAATTGACT TTCATTTGGA 840

AAAAGCCCCA CTTCACTGTG GTCTGTCACT TTGATGGTGT CAGAGGGTCC AGGACCTCAG 900

TGCCAGGGTG CGAGGAGAGC AGTGCTGTGC AGTGGGGAGG AACCTCACCA TCACCCAGTC 960

TCCTCGCCAG AGGGTCCAGG ACCTCAGTAC CGGGGTGCGA GGAGAGCAGC GCTGTCCAGC 1020

GGGGAGGAGT CTCACCATCA CCCAGTCTCC TCACCGTCAC CCAGTCTCCT CGCCAGAGGG 1080

TCCAGGACCT CAGTGCCGGG GTGCGAGGAG AGCAGTGCTG TCCAGCGGGG AGGAACCTCA 1140

CCATCACTCA GTCTCCTCAC CAGCACACTT TTTCTCCATG TCTCGTTTTG CTCCTCCTCT 1200

GGTATTTGTA TTTCTTAAAG AGGATTTTGA AAAGAGATGG TGAAGTTGGT ATTTTAGGTA 1260

GAAGGGACCA ATTGTTTCCT CAAGACTAAG TTGGTCCAAC CAAACTGACA GAGACGAGGT 1320

CTCTACATAT GAAAGATGGA ACCTGGCCGG GTGCTTCGGG GGCTCGCGCC TGTAATCCCA 1380

GCACTTTGGG AGGCCGAGAC GGATGGATCA CTTGAGGTCA GGAGTTTGAG ACCAGCCTGG 1440

GCAACATAGC GAGACTCCAT CTCTACAAAA AATAAACAAA ATTAGGCTGG TGTGGTGGCG 1500

AGTGTCTGTA GTCCTGTCTA CTCAGGAGGC TGAGGTGGAA GGATCACCCG AGCCCAAGAG 1560

GTCAGGGCTG CAGTAAGCCA TGGTCACGCC ACTGCACTCC AGCCTGGGCC ACAGAATGAG 1620

ATCCCGCCTG TCTCTTACAA AAAAAAAAAA AAAAAA 1656

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1151 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

CACCCACCTC AGCCTCCGAA GTACCTGGGA CTGTAGGCAC AAGTCATGCC TGACCATGCC 60

AGCTCATTTT TTATTTATTT TAATTTTTTT TGTAGAGATG GTGTCTTGCT GTGTTACCCA 120

GGCTAGTCTC GAGTTTCTGG TCTCAAGTGA TTTTCCAGCC TTGGTTTCCC GAAATGCTGG 180

ATTACAGGCG TGAGCCACCA TGCCCAGTTT AAATAGTAAT CTGTAAAGAA CAGCTAGCAC 240

TCTCATGAGT GTTCCATGTT GAGACTCTGT TCTCAGCACT GTATATACTG ACTCATGTGA 300

TCCTCATAAT AAGGCACAAA GAAGGGGCAG TTATTCGTAC AGATGAGGAA AATGAGGCAT 360

AGAAAAGTTT GGTAACTTGC CCAAGGTCAC ACAGCTTGTT TGTAGCAGAA TCCGGATAAG 420

GCTTGTGCAC TGAGGTGGCA TTTGCAGCTT CCCTGAGAGG GCCCTCTGCA CACATCATCT 480

CTGATCCTCA GACAACCCTG CAGAGAGGTG GGAGGTGTTG TAAGCTCCAT TCCTCCCCAA 540

ACTGGCATCA CCCAGCAAGC TGGGATTCAG ACCAAGGGTG CCAGACTCCA GAACCCGTGG 600

TCTTGTCTCT GCACCTCAGT GCCCGTCCCC CGCCATGGTC TGGCTTCCTT TCCTTTCTCC 660

TCCAAGTCTC CTTCTCACTT TGCTACCATC TTTGCTCTGA GCAGCTGCTG ACGACCCAGC 720

GGGTGAGCTG CGCCCACATC TACAGTGCCC TAGACCCGAC AGCCCGCAAG ATCAATCTCG 780

CCAAATTCAC GCTTGGCAAG TGCTCCACTC TCATTGTGAC TGACCTGGCC GCCCGAGGCC 840

TGGACATCCC GCTGCTGGAC AATGTCATCA ACTACAGCTT CCCCGCCAAG GGCAAACTCT 900

TCCTGCACCG CGTGGGTAAG CAGCCCGTGG CTGGCCCTGG GGCAGGCAGG GGTGCCGGAT 960

CCTGGCAGAA GCCGAGGGTA CAAGGCTTAA CTCTTGACAC TGCACATGGG GTGGCTGTGG 1020

GCCTTGTTTT AGAGACAGAG CCTCGCTATA TTGCTTAAGC TGGTCTCAAA CTCGCGATCG 1080

TGCCACTGCG ATCCAGCCTG GGTGACGGAG CAAGATCTTG TCTCAAAAAA ACCAAAAAAA 1140

AAAAAAAAA A 1151

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1299 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GACAGTTTGC TTATCCATTC ATAAATTGAT GGACATTTGG GTTGTTTTCA CTTTTTGGCT 60

ATTATAAATA ATGCTGCTAT GAATACTCAA GTATGAGATT TTGTGTGAAC ATGTTTTTAG 120

TTCTCTTGTG TATACTGAAG AGTTTAATTA TTGGTCATGT GATGATAACT CTAAATTTAA 180

CTTTTTGACG AACTGTCAAA CTGTTTTCCA AAGTGACTAT ACTATTTTAT ATTCCCACCA 240

GCAGTGAATA AACATTCCAA TTTTGCCATA CTCACCAACC TTGTACTTGT CCAAGCCATC 300

ATAGTGGGTA TAAAAGTATT TCCTTGTGGT TCTGGCTATG CCCTAATGAC TGTGAGGCTG 360

AACATCTTTT CAAGTGTGAA TTGGCCATTT ATATACCTTC TTTGGAGAAC TGTCTTTTCA 420

AACCCTTTGC TCCTTTTTAC ATTGAGTTAT CCATCTTTTA ATTGTTGGGT TGTATATTGT 480

TTAATTTGAA AATCCATGTT ATGTATAATA TGTGTAATTC TAAAATTGTT TATTCTTACC 540

AAGTTGCCAG CTATCAGAAC ACTAATTTGT TGCATTATTT TTCCCCTTTA ACATTAGTTT 600

GTTCTGCTTC CTTTATTAAT AATTAATAAT GGGCTGGGTG CCGTGCCTCA CACCTGTAAT 660

ATCAGCACTT TGGGAGGCCG AGGCAGTGGA TCATTTGAGG TCAGGAAGTT CGAGACCAGC 720

CTGGCCAACA TGGTGAAACC TCGTCTCTAC TAAAAATACA AACATTAGCT AGGTGTGGTG 780

GTGCATGCCT GTAATCCCAG CTACTTGGGA GGCGGAGGCA GGAGAATTGC TTGAGCCTGG 840

GAGACGGAGG TTGCAGTGAG CCGAGATCAT GCCACTGTAC TCCAGTCTTG GCGACAGAGT 900

GAGACCCAGT CTCAAAAAAT AGTAATAATA ATGTATTAGT TTGTGCTGCT GCTTTATCAA 960

ATAACTTATT CTTATAAAAT ACATAAGAGG GTTGAGTGTG GTGGCTCACG CCTGTAATCC 1020

CAGCACTTTG GGAGGCTGAG GAGCATGGAT CACTTGAGGT CAGAAGTTCG AGACCAGCNT 1080

GGCCAACCTG GCAAAACCCC ATCTNTACTA AAAATAGAAA AAAATTGGCC AGGCATGGTG 1140

GCACGTGCCT GTAGTCCCAG CTANTCAGGA GGCTGAAGCA GGAGAATTGC TTGAATNTGG 1200

GAGGCGGAGG TTGCAGTGAG TCGAGATAGC ACTCACTGCA CTCCAGCCTG GGTGACAGAG 1260

CAAGACTCAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAA 1299

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 871 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GGCASRRNAG ACAGACCTGA GTGCCAASGK TGTGACCTCA GGCCTCTCCA GGTCTCAGTT 60

TCCACATCCG TGAAATGGGT GTGATGAGAG GGTGACGAGG AGGGGCCAGG ACGGGGAGGC 120

CACGGGGAGG CCAAGGGGTT GGGGCAGGAC TGGTCACAGT GGCTCCAAGT GCCCATTCAG 180

GCAGTARGCA ATGGGGTTGA GGTCCCTGAA CTCTCTCTCC AGTGTGATGT TCTTGGTGCA 240

TGGGGGTGCC TGGGGTGCTC CCAAGGCCTC CGCCCGCCAC CTCTGTCTCT CCCTGGGCCT 300

CCATCCTTCC ACCTGGCTCT GGAATCACAA CCGGTGGTAG CCAGTCCCCA GGACAGCTTC 360

CAGTCCCTTA GATAGTCACC CTCATGAGCC CACCCAGCCT CTGGGTTGAC ATAAACACCC 420

CCAGCAGCCC CTARCTGCCT CTGGCTGACA TCAACTGTAK GACATGGGGC CTGGAACCTG 480

GGAAACAGCT ACCTCGGGGG GAATGCTGTT GGTGAGGGCC AGGCTCTGGG TTCCCATCCC 540

AGCTGCTTAC TAAGAATCAT GGGGTGTGTA GGCCGGGTGT GGTGGCTCAC ATCTATAATC 600

CCAGCACTTT GGGAGGCTAA GGTGAGTGGA TCACCTGAGG TCAGGAGTTC GAGACCAGCC 660

TGGCCAACAT GGTAAAACCC CCTCTCTACT AAAAATACAA AAATTAGTGG GCATGGTGGT 720

GGGCGCCTGT AATCCCAGCT ACTCGGGAGG CCAAGGCAGG AGAATCACTT GAACCCGGGA 780

GGTGGAGGTA GCAGTGAGCT GAGATTGAGC CATTGCACTC CAGCCTGAGT AACAGAGTGA 840

GACTCCGTCT CAAAAAAAAA AAAANGNAAA A 871

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1023 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GTCACCAGGC CCCCACTCAA TCTCAGCTTG GGAACCAAAG TCACCCACCT TGGTTGTGTT 60

GGGGTGGACC CGCCATCTGT CCCTGGTCCA GACGAGAAAG AGGAGTCTCT CCTAGGCCTG 120

GAGCTGGGAA AGAATGTGTG CCCCAGTTGT CTGCACTTCT AATTCTCATC ATGGAGAAAC 180

CTTTATTCCT ATCTCCCTTT CCTGAACTGG TTTTTTGTTG TTTTTGTTTC ATTTTGTTTT 240

GGGGGGATAG TTTCTTGCTC TTTAATTTGG AGTCTCCAGT ACCTTTGGGA TGCAGGCAGT 300

TCTTGCCTGG GCCTTCTCGG AACCCTCACT CCCCTAGCCC ACTCTTGCGC TACCTGCAGG 360

AGGCTGCCAA CCTGGTGCAT TCTGACAAGC CTCCCACCCA AATCTCTCTC CTGCCATTGT 420

GTCCAAAATC CCACCATTAG ATGCTCTTGT AGGGAAGAGC GTTTCTTGAA GGCTTTTAGG 480

CCTTCCAGAG CCAGGAGGGA AGTCAGACAA TAGCAGGAAG TCCCCAGGCC TTTTCAAAGT 540

TCCAAACCAA GCTCTCCTGA TTTTAATGTA GAGATCATAC CAACCCAGGT GGGGGAGGAG 600

GGTCCCCAGC CCCAGGCAGC AGCCATCACC CCCTCCACTG AAAACAATAT TGGAGGCTGC 660

TTTGGGACTG CCCTTCTCAG CCCCCTAAGT CTGTTTTGTA ATGCCTGTGG TGCTCTCCCT 720

CCTGGACCTT TCCTCTCGGG GGTCACCACA CTTTGCTAAC TCTTGTGTGC ACATATTTTA 780

TAATAGAGTA GCGAGGGAAT GGTGCCGCCT CCAGCTTCCG TAAGCTGCCC GGGCTCTGGG 840

GGGCTCTGGG ACAATCGGGG CTGGGAAGTG ACTGTGCTCT TATTGTACAC TCTTTATTTC 900

TCTGTATCTT TGGCTTGTGC TCTTTGTAAT TAATGGGATT TGTCTGCCTT TTCAACACTA 960

TACTGAGCAA TAACAATAAA TGCACACGTG GAAATGCAAA AAAAAAAAAA AAAAAAACT 1020

CGA 1023

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1085 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CCGTTTTGAA GGTCCTAGCC CACCTGGTNN GNCTCACGCG CACGACTAGC CGCTCCCATA 60

CAGCACGCCC GGACTCTGTC GTCGCTTAAG GCCACTCCTA TTCTACGGCT GACCCCTGGT 120

GGTCACGTGG ATCTGTTCGC CACGCAAGTC TGGGTCCTTC GGCGATTGAC CGGGGTCCTT 180

GCTGTTCGGG AGCCTCTCCT AAGCTGCCTG TTCGCGCGAR AKTTTGGAGG GGCGGGTTTG 240

GGGTCGGTGT CTGATTGGGG CTCGCACCGC AGCACGCTGG AGTCCCGCTT AGGTACCAGT 300

TAGCGTCAGG GGAGCTGGGT CAGGCGGTCG CGGGACACCC CGTGTGTGGC AGGCGGCGAA 360

NGCTCTGGAG AATCCCGGAC AGCCCTGCTC CCTGCAGCCA GGTGTAGTTT CGGGAGCCAC 420

TGGGGCCAAA GTGAGAGTCC AGCGGTCTTC CAGCGCTTGG GCCACGGCGG CGGCCCTGGG 480

AGCAGAGGTG GAGCGACCCC ATTACGCTAA AGATGAAAGG CTGGGGTTGG CTGGCCCTGC 540

TTCTGGGGGC CCTGCTGGGA ACCGCCTGGG CTCGGAGGAG CCAGGATCTC CACTGTGGAG 600

CATGCAGGGC TCTGGTGGAT GAACTAGAAT GGGAAATTGC CCAGGTGGAC CCCAAGAAGA 660

CCATTCAGAT GGGATCTTTC CGGATCAATC CAGATGGCAG CCAKYCAGTG GTGGARAATT 720

GTWCAATGAN GTGCCGGGCA CGGTGGTTCA TGCCTGTGGT CCCAGCACTT TGGGAGGCTG 780

AGGCGGGTGG ATTACCTGAA GTTGGGAGTT TGAGACCAGC CTGACCAACA TGGAGAAACC 840

CCGTCTNTAC TAAAAATACA AAATTAGCTG GGCGTGGTGG CACATGCCTG TGATCCCAAC 900

TACTCGGGAA GCTGAGGCAG GAGAATCACT TGAACCCGGG AGGTGGAGCT TGCGGTGAGC 960

CGAGATCGCG CCATTGCACT CCAGCCTGGG CAACAAGAGT GAAACTCCAT CTCAAAAAAA 1020

AAAAGAAAAA AAAAAAGAAT TGTACAATGA GGTAAAATAA AATCATATAG TTGAAACTAA 1080

AAAAA 1085

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1361 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GTTATTGTTT TTTTTTTTTT CTACTACTAA TTCCTAGTGG TGCACAACGA GTTTCTGAGA 60

ACACAGTAAT AACGCCAAAG TAGCAGTGCA TCTGAGAAAC ATAATTTTTA CCTCGTTGCT 120

TTCCCAAAAA TAAATATCTC AGTCATGGAA AACACTGTTT ATTTGAAAAC AATGAGACCT 180

CAAATATGAA ATATAGTTAA CAATGACATT GACACTGTTG CTAGCACTTT CCCCTAAACC 240

ACCCGTAAGT CTTGGACGCA TGTGCATGCA GCACACACAC ACACACACAA AAACCAAAAA 300

CAAAGCCAAA AAAAAAAAAA TCCCAAACAC AACAWTCCAT GATTGTTCAA TGACTCCTGA 360

TGCCGGGAGG ACAGGCTGTT AAAAGAATTT GTCTCCCACA ATATCTCTGG AGTGGGCACA 420

AAGCCCATCA CCTGTTAGTG ATCACAGACA TTCAGTTAAC CTGTCCTTCC AGTAATCAGA 480

GACAACAATT CAGACCCTGG ACTTCTCAGA ATCCATGTAC TGCTGAGTCT TGGCTTTGAG 540

ACAAGACAAG TCTTGGCTAA ATTGAGGCAG GACAGCAGCC CCTTCCATAT GTTTGGTCCC 600

ATTTGATAGA AAGTCTAATT TAGAGTTATA AATGTGCTCA TCTATTTACT CTGAGCTCAA 660

TCTAATTTGA CAGGTAATTC CTCACATTTT CTCCATTAGC CAGCTGAGAG TCAGCTGTGG 720

TAGAGACACA CGACATGGGT TCAAGCCCCT CATGAGCCCT GTGGTGGCTG GCAAGTCCTT 780

TCCTTTCTTT AAGCCTTAAT CTCCTCACTT GATAGAGGGG GAGAAATTGA CCCAATGATG 840

ATAAATATTG TGTGGTTCTA TATTTCTAGC CTAGACAATT GTTGCTCAAG TGTAACATGT 900

GACTGCCAAA TAGGATATCT CTTAAGATGA ATATCTCCTA ACTTTCCTCA CCTGGTATGA 960

TCACATATTC TGGCTTCCTC TAAGGTTTAG AATCTGTAGG TTCAAAAAGG TCTTGAAGAC 1020

CACCTAGCAT ATTTCTATTC TGATGAGAGA AACTCCATCC AAGCAGCTGT ACTTTTTCAA 1080

CTTGAAAATC TCCAAGGAAA AGAGCCTTTT CCTTCTGTTG ACTAATGTTT AGATGCAGAC 1140

CAGGGAATCT TTCCTGTCAC TTGGCTKWWY TCCCTTGCTA AAACAACATA AGACAAGTAT 1200

YWKWTCCCCT TTATTGGAGG ATCCAAGGGA AGATYAGCTG GTCACTAAAG TCCAGAAAGC 1260

AATGGAGTCT TATAGCTCAT TCCTGGGATG TTGCAAATAA TGCCAAACTC TGATGTACTC 1320

AGACTCAACT TCTAGGATTA TCATTCACTA AATGCCTGGG T 1361

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1822 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

TCATTTGGTT GCCAAATSST AGATTAGCGC TGGGTCAGCT GTGGAAGAAT CGCCTCAGCT 60

TACAGATTGT CAGACAGATC TAASTAGTTT TCCAGAAAGC CTGGGAAGCT GTGTGTTCAA 120

CATTTCCCAA GGGATTCTGA TCACCAGGCA GCTTGGGAAC CACTGGGGCA GGCCAAATAG 180

AATATTTTGG GCGGGAAAGA AGCACCCGAT TTAAAATGAA GCGTAACCAG AGGAGTTCAG 240

AACTGGGAAG AGAGTGGTAG ACTTCCTGTG ATCTTCAGAA ATCATCTACC TGGTAAAAAT 300

ACATGCTGTT TAGAATATCT GATAGGTGTT TCCAGCTACT ATTAGAGGTG ATAGTGCTTT 360

TGTGGGGGAA AAAATTGGTC ATGGTGAATG GAGATCGAGG AAGCTCGGGA CAAGGGAGGG 420

GTGGGCTGCC TGATTTTGTC CAGTTTTCCA AATATCCACG CAGAANCTGG AGTATCCTAA 480

ACATGAGAAT GTACAGTTGA CAGTTGTAAA AACTAGGGAT CTGTAGTGAA TGCTGTGCAG 540

CCCCATATCT CATTTGGGGG TAGGAAAATA GCTGAAGATT CATGTGCATT ATTTGACATT 600

TCCTTTGTCA TCTGCTTTTT AAGCAAAAAA GGGTTTTGTG TTAGAAATTC TACTTGAGCA 660

GATTATAAAG AGCTTTAAAA AACAACTTTC GGTTGCCAAA AGTTTGAGCA TTTGATTTCA 720

TTACCTGTGT CTCCCTCACT GGTGTCCAGA CGGTCAACTG AATACTCCTG AAACCCAGGG 780

AGCAGGTGAC TTCCTGGAGT GCTTTGTCCC CAGAGTCAGC CACTGCTTCC TCTGTGGGGG 840

TGGAGAGTTT GTCTTTGGCC ATGCAGTGTG CGACAGTTCA GGACGGGTAG GGATGGGTCC 900

CATTCTGTCT GGGTCAAGGG CTCTATCAGC TTCTTCCATG TGCCTTTGGG AAGAAATCTC 960

GTTACTTTAA GTTTGCTTTC CTGTTATCTT GATGAAGTGC CCATTTTAGC AGACACTTGT 1020

AGTGCTGACC ACTTAGGGAA TGTACAAACT CCTAAGCTTC TAAAGGGAGG CATGGCAAAA 1080

ACGTTGGGGT CAGGATGTCT CTCACGCTGC TCATGTTAAT ACTATTAACA CATGATTTGA 1140

GAAATAAGTT TTCTCTAAAA TGCATATTTT GCCGCCACAC ACTGAACAAT ATTATTTCCA 1200

GTGAAGTTTG ATGCCTGTTC TTACGTTGTG TTCACCTGTT GGTTCACCAC TCAGCAGATC 1260

TGATTCTGCA AGAATTAATG GTAGAACTAG ATCATCCTTT CTAACAGACG AGCCTGTGTC 1320

CTGTGACGGC CTTTCACAGC GGAATGCAGT TGTACCTCAC ATTACTTTTG AAACTTCACT 1380

CGTTCCAGTT GGTACAAGTA TTTGCCAAAG CCATTTCCTA TGTTCACCGT GGCCCCTCCT 1440

GATGTGGCTG TCAGCGCAGC GTTGNTTGAA CAGGGCTATT CTTTTTACAA GGTGTGAAGT 1500

GTGGCTCTTC GCTTCGTCTT TGCCATGGCA TTAAAAGAAA GTTCCCTGTC TTCTTTCAAT 1560

ATTAGTTATT TCAAATGAAT ATGTGCTACT TAAAAGCTTG TTTTGTTTCT TTGTATATAA 1620

TTTGCCTTGG ATTTATTGTG CACAGTTTGT TGAGTTGTAT GTTTTTGTGA ATTATCAGGA 1680

GTAAATTTGA CAAGTACATG TGAATAACCT CCTGTAAATG AATTTTATAA CAAAAATGTA 1740

CTGAACTATT TTTTAAAGTT GTGCAGATTA GCAAAAAAAA AAAAAAAAAA AAACTCGAGG 1800

GGGGCCCCGT ACCCTTTTCG AA 1822

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1873 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TCTAAATAAA AGGGTCTAAA ACTCAGCTTC TGAGTTTTTA AAATCACGGT CTCCAGGTAC 60

CAATAAATGC TACAGTTTGC CTTATGATGT TAACATAAAA CACTTAGTAG AAGGACAATA 120

TTTCCATGAA AATAATGTTT TTCAATATTA AGAAGTTACT ACTCAAATTT TCACAGTAAG 180

CCATTTAGGG TATGTTTGGC TATTTTTATA AGGACATGAG AGATTATGTC ATAATTTTGT 240

TGTGGAAGTC TCACTCTTGG CTAACTTAAA AGCATTGTGG ATAGTAGCAG TTACTAGTTC 300

CAGGTTGTCA TATTTACAGG AAAATATGTA TATGGTGAAA GGCCACCGTG TTTAATTACT 360

ATAATGATGT AGAAAAGATT CCCGTGTGAA TTTTTTTTTT GAAAGTCTAA AAAATGTATG 420

CTGTAAAAAT TTGCTGCAGT GTAATTTTGC ATTCTCTTTA AACTGATTGA GGTCACAGTA 480

TTTTATTATT TGGGGTCCTC ACCACAGGAA ACACTGCGAT ACAGGGGCAA AAGAGATGGC 540

AGTGCAATTT AAATTAATAC AACAAAATCA ATGCAGCACC AACCAAGACT GCCAGGTCTG 600

GTGTCATGGG TATGCCCAGA GCCCAGGAGT TCAGAAGGGC CCTAAGCCTG ATTTAATGCT 660

CTGCTGTTGA TGTCTTGAAA TTCTTAACAA TTTTTGAACA AGGGGCCTGC GTTTTCACTT 720

CGCACTGGGC CTTGCAAATT ACATAGCGAG TGCTCATAAA AGAACTCAGA AACGTGGTAC 780

CTCTCTTCCT GGTGGATACA AATAAAGAAA TCTGGATCCA AAGTTGAAAG TTGCTGGCGA 840

TATCATTCAA GTAGGACTCT AAATAGTGGA TTAAGATGAG GGTGGGCCTG GGTGAAGATT 900

CTTTCCAGCT TTAAAAGAAA GTGACTTCAA AAACTGACTG CAAATATTGA CGATGGTTTC 960

TGCTGGAGGA AAAGAAACAG CTTGAATACA GACAGGCTTT TTTATTACGG TACTGATATA 1020

TTGACCTTAA ACTTGCTGAG GAACTGAACT AACGTCCTCC AGTGACCGTG GAATTCCATC 1080

TCAGCTCCAG GAACATGCAG ATACCTGCAA AGAGACACGC ATATATGCTG GCATACATGT 1140

GCATTTGGTG TTGGGAAGTT GACCATCTGG TCTATCTTAA TAAAATGGTA AAAAGCACAC 1200

CAAGACAATG ATGGGGGCAG GAGGATGTTT TTGAAAACAG CGCTTCTCAA CCAGTGCTCG 1260

ATTTTGCCCC CCAGGAGACA TTTGGCAATG CAATGGCAAC TTTTGGTTGT CGCAGCCGGG 1320

GAAGGGAAGC TACCAGCATC TAGTGCGTAG AGGTCATGGA CGCCGTTAAA CATCCTACAG 1380

TGCAAGCGCA SCCCCNGACC ACGAAGAGTT GTCTTGCTCA AATATCAACA GTGCTGCAGT 1440

GTAGAAACTT GATCGTTGGT TTTCTTTTAA TGCAAAACTC TCATAAAAAC CTTTCACTTT 1500

TCCTGTCATT GATTATATGC TTGATACACC CAAAAAGAAA AGGGGAGGGG CACCAATTCA 1560

CCTACACTCC AGTGGCTCCA TCACCTTTAA AAATATTTAT AAAATAGTTC CAAAAATCTG 1620

ATATCTGAAA AGCAATCCAA GCCTGTGTAA ATGGGAATCA CTGATAAGTA TCATCATCTG 1680

TATCAGCTTG GCTTGGACAT GAAAAATTGA TTCTCTTTAT GTCACTCCTT GCACCTGGAC 1740

AAATTCAATC CCCGGTACTT AAGTCACACT GCCAASCCTC GGCCCTGACT ATTGTCTTGA 1800

TTGCTGTTCC TTTCTGGTTC AAAATAAAAT CATTTTTGTG GCACCAAGAA AAAAAAAAAA 1860

AAAAAAAACT CGA 1873

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 865 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GGCACGAGAG CAGACCTGAG TGCCAACGGT GTGACCTCAG GCCTCTCCAG GTCTCAGTTT 60

CCACATCCGT GTAAATGGGT GTGATGAGAG GGTGACGAAG AAGGGCCATG ACGGGGAGGC 120

CACGGGGAAG CCAAGGGGTT GGGGCAGGAC TGGTCACAGT GGCTCCAAGT GCCCATTCAG 180

GCAGTAGCAA TGGGGTTGAG GTCCCTAACT CTCTCTCCAG TGTGATGTTC TTGGTGCATG 240

GGGGTGCCTG GGGTGCTCCC AAGGCCTCCG CCCGCCACCT CTGTCTCTCC CTGGGCCTCC 300

ATCCTTCCAC CTGGCTCTGG AATCACAACC GGTGGTAGCC AGTCCCCAGG ACAGCTCCAG 360

TCCCTTAGAT AGTCACCCTC ATGAGCCCAC CCAGCCTCTG GGTTGACATA CACACCCCCA 420

GCAGCCCCTA GCTGCCTCTG GCTGACATCA ACTGAGGACA TGGGGCCTGG AACCTGGGAA 480

ACAGCTACCT CGGGGAATGC TGTTGGTGAG GGCCAGGCTC TGGGTTCCCA TCCCAGCTGC 540

TTACTAAGAA TCATGGGGTG TGTAGGCCGG GTGTGGTGGC TCACATCTAT AATCCCAGCA 600

CTTTGGGAGG CTAAGGTGAG TGGATCACCT GAGGTCAGGA GTTCGAGACC AGCCTGGCCA 660

ACATGGTAAA ACCCCCTCTC TACTAAAAAT ACAAAAATTA GTGGGCATGG TGGTGGGCGC 720

CTGTAATCCC AGCTACTCGG GAGGCCAAGG CAGGAGAATC ACTTGAACCC GGGAGGTGGA 780

GGTAGCAGTG AGCTGAGATT GAGCCATTGC ACTCCAGCCT GAGTAACAGA GTGAGACTCC 840

GTCTCAAAAA AAAAAAAAAA AAAAA 865

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

Met Gly Arg Asn Ile Leu Ile Ile Thr Val Val Thr Cys Val Asp Leu
1 5 10 15

Arg Pro Ser Ser Met Ser Ser Leu Ser Ala Thr Cys His Ser Thr Trp
20 25 30

Thr Arg Ser Ser Gly Cys Phe Xaa Ser Ala Ala Leu Pro Ala Thr Ser
35 40 45

Ser Pro Trp Arg Lys Gln Arg Xaa
50 55

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 183 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

```
Met Lys Gly Trp Gly Trp Leu Ala Leu Leu Leu Gly Ala Leu Leu Gly
 1               5                  10                  15

Thr Ala Trp Ala Arg Arg Ser Gln Asp Leu His Cys Gly Ala Cys Arg
            20                  25                  30

Ala Leu Val Asp Glu Leu Glu Trp Glu Ile Ala Gln Val Asp Pro Lys
        35                  40                  45

Lys Thr Ile Gln Met Gly Ser Phe Arg Ile Asn Pro Asp Gly Ser Gln
    50                  55                  60

Ser Val Val Glu Val Pro Tyr Ala Arg Ser Glu Ala His Leu Thr Glu
65                  70                  75                  80

Leu Leu Glu Glu Ile Cys Asp Arg Met Lys Glu Tyr Gly Glu Gln Ile
                85                  90                  95

Asp Pro Ser Thr His Arg Lys Asn Tyr Val Arg Val Val Gly Arg Asn
            100                 105                 110

Gly Glu Ser Ser Glu Leu Asp Leu Gln Gly Ile Arg Ile Asp Ser Asp
        115                 120                 125

Ile Ser Gly Thr Leu Lys Phe Ala Cys Glu Ser Ile Val Glu Glu Tyr
    130                 135                 140

Glu Asp Glu Leu Ile Glu Phe Phe Ser Arg Glu Ala Asp Asn Val Lys
145                 150                 155                 160

Asp Lys Leu Cys Ser Lys Arg Thr Asp Leu Cys Asp His Ala Leu His
                165                 170                 175

Ile Ser His Asp Glu Leu Xaa
            180
```

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

```
Met Phe Thr Leu Ala Phe Phe Phe Leu Ile Asn Phe Leu Asn Val Lys
 1               5                  10                  15

Tyr Asp Arg Xaa Ser Xaa
            20
```

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

```
Met Gly Phe Val Pro Thr Pro Glu Ile Leu Trp Glu Thr Asn Ser Phe
 1               5                  10                  15

Asn Ser Leu Ser Ser Arg His Gln Glu Ser Leu Asn Asn His Gly Leu
            20                  25                  30

Leu Cys Leu Gly Phe Phe Phe Phe Leu Ala Leu Phe Leu Val Phe Val
        35                  40                  45

Cys Val Cys Val Cys Cys Met His Met Arg Pro Arg Leu Thr Gly Gly
    50                  55                  60

Leu Gly Glu Ser Ala Ser Asn Ser Val Asn Val Ile Val Asn Tyr Ile
65                  70                  75                  80

Ser Tyr Leu Arg Ser His Cys Phe Gln Ile Asn Ser Val Phe His Glu
                85                  90                  95

Lys Lys Lys Lys Lys Asn Ser Cys Gly Arg Gln
            100                 105
```

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

```
Met Phe Leu Val Thr Pro Ala Thr Leu Trp Ser Val Pro Cys Phe Leu
 1               5                  10                  15

Leu His Ser Trp Pro Pro Ser Pro Ala Pro His Thr Gln Met Leu Ser
            20                  25                  30

Leu Arg Glu Ala Gly Thr Ala Trp Gln Ser Glu Lys Ser Val Ser Xaa
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

```
Met Lys Thr Ser Ala Leu Leu Pro Phe Ser Ser Ser Gln Gln Pro Gly
 1               5                  10                  15

Ile Leu Lys Pro Xaa Gly Ala Gly Thr Cys Asn Ala Gln Glu Pro Ser
            20                  25                  30
```

```
Xaa His Leu Glu Ser Thr Ser Asp Pro Arg Trp Gly Gly Pro Cys Arg
        35                  40                  45

Pro Ala Val Pro Gly Gly Leu Ser Met Ala Val Trp Lys Ala Trp Val
     50                  55                  60

Ala Gly Met Trp Leu Ser Leu Pro Pro Leu Asn Leu Arg Ser Cys Trp
 65                  70                  75                  80

Glu Thr Xaa
```

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 315 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

```
Met Val Lys Leu Leu Val Ala Lys Ile Leu Cys Met Val Gly Val Phe
  1               5                  10                  15

Phe Phe Met Leu Leu Gly Ser Leu Leu Pro Val Lys Ile Ile Glu Thr
             20                  25                  30

Asp Phe Glu Lys Ala His Arg Ser Lys Lys Ile Leu Ser Leu Cys Asn
         35                  40                  45

Thr Phe Gly Gly Gly Val Phe Leu Ala Thr Cys Phe Asn Ala Leu Leu
     50                  55                  60

Pro Ala Val Arg Glu Lys Leu Gln Lys Val Leu Ser Leu Gly His Ile
 65                  70                  75                  80

Ser Thr Asp Tyr Pro Leu Ala Glu Thr Ile Leu Leu Leu Gly Phe Phe
                 85                  90                  95

Met Thr Val Phe Leu Glu Gln Leu Ile Leu Thr Phe Arg Lys Glu Lys
            100                 105                 110

Pro Ser Phe Ile Asp Leu Glu Thr Phe Asn Ala Gly Ser Asp Val Gly
        115                 120                 125

Ser Asp Ser Glu Tyr Glu Ser Pro Phe Met Gly Gly Ala Arg Gly His
    130                 135                 140

Ala Leu Tyr Val Glu Pro His Gly His Gly Pro Ser Leu Ser Val Gln
145                 150                 155                 160

Gly Leu Ser Arg Ala Ser Pro Val Arg Leu Leu Ser Leu Ala Phe Ala
                165                 170                 175

Leu Ser Ala His Ser Val Phe Glu Gly Leu Ala Leu Gly Leu Gln Glu
            180                 185                 190

Glu Gly Glu Lys Val Val Ser Leu Phe Val Gly Val Ala Val His Glu
```

```
        195                 200                 205

Thr Leu Val Ala Val Ala Leu Gly Ile Ser Met Ala Arg Ser Ala Met
    210                 215                 220

Pro Leu Arg Asp Ala Ala Lys Leu Ala Val Thr Val Ser Ala Met Ile
225                 230                 235                 240

Pro Leu Gly Ile Gly Leu Gly Leu Gly Ile Glu Ser Ala Gln Gly Val
                245                 250                 255

Pro Gly Ser Val Ala Ser Val Leu Leu Gln Gly Leu Ala Gly Gly Thr
            260                 265                 270

Phe Leu Phe Ile Thr Phe Leu Glu Ile Leu Ala Lys Glu Leu Glu Glu
        275                 280                 285

Lys Ser Asp Arg Leu Leu Lys Val Leu Phe Leu Val Leu Gly Xaa Thr
    290                 295                 300

Val Leu Ala Gly Met Val Phe Leu Lys Trp Xaa
305                 310                 315
```

(2) INFORMATION FOR SEQ ID NO: 42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

```
Met His Met Asn Leu Gln Leu Phe Ser Tyr Pro Gln Met Arg Tyr Gly
  1               5                  10                  15

Ala Ala Gln His Ser Leu Gln Ile Pro Ser Phe Tyr Asn Cys Gln Leu
             20                  25                  30

Tyr Ile Leu Met Phe Arg Ile Leu Gln Val Xaa Ala Trp Ile Phe Gly
         35                  40                  45

Lys Leu Asp Lys Ile Arg Gln Pro Thr Pro Pro Leu Ser Arg Ala Ser
     50                  55                  60

Ser Ile Ser Ile His His Asp Gln Phe Phe Pro Pro Gln Lys His Tyr
 65                  70                  75                  80

His Leu
```

(2) INFORMATION FOR SEQ ID NO: 43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```
Met Gln Asn Ser His Lys Asn Leu Ser Leu Phe Leu Ser Leu Ile Ile
 1               5                  10                  15

Cys Leu Ile His Pro Lys Arg Lys Gly Glu Gly His Gln Phe Thr Tyr
            20                  25                  30

Thr Pro Val Ala Pro Ser Pro Leu Lys Ile Phe Ile Lys
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

```
Met Leu Arg Lys Tyr Met Pro Glu Thr Ser Val His Cys Leu Ala Leu
 1               5                  10                  15

Thr Val Leu Val Glu Thr His Ser Gln Thr Lys Pro Thr Ala Ala Phe
            20                  25                  30

Leu Trp Ser Gln Phe Met Phe Leu Ile Leu Ser Phe Gln Xaa
        35                  40                  45
```

(2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

```
Met Lys Leu Ala Leu Phe Pro Leu Phe Cys Phe Ser Arg Ile Leu Arg
 1               5                  10                  15

Lys Ser Thr Asp Xaa
            20
```

(2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 43 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

```
Met Gly Ser Ser Glu Thr Ser Leu Leu Gly Leu Gln Leu Val Thr Phe
 1               5                  10                  15

Leu Leu Leu His Met Val Leu Leu Leu Cys Val Thr Val Ser Lys Phe
            20                  25                  30
```

Pro Phe Cys Lys Asp Thr Ala Ile Leu Asp Xaa
35 40

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

Met Arg Pro Gly Leu Ser Phe Leu Leu Ala Leu Leu Phe Phe Leu Gly
1 5 10 15

Gln Ala Ala Gly Asp Leu Gly Asp Val Gly Pro Pro Ile Pro Ser Pro
20 25 30

Gly Phe Ser Ser Phe Pro Gly Val Asp Ser Ser Ser Ser Phe Ser Ser
35 40 45

Ser Ser Arg Ser Gly Ser Ser Ser Ser Arg Ser Leu Gly Ser Gly Gly
50 55 60

Ser Val Ser Gln Leu Phe Ser Asn Phe Thr Gly Ser Val Asp Asp Arg
65 70 75 80

Gly Thr Cys Gln Cys Ser Val Ser Leu Pro Asp Thr Xaa Phe Pro Val
85 90 95

Asp Arg Val Glu Arg Leu Gly Ile His Ser Ser Cys Ser Phe Ser Glu
100 105 110

Val Xaa

(2) INFORMATION FOR SEQ ID NO: 48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

Gly Arg Lys Gly Gly Leu Ser Gly Thr Ser Phe Phe Thr Trp Phe Met
1 5 10 15

Val Ile Ala Leu Leu Gly Val Trp Thr Ser Val Pro Val Val Trp Phe
20 25 30

Asp Leu Val Val Asp Glu Gln Ile Thr Ser Gln Ser Lys Gly Leu Pro
35 40 45

Leu Xaa
50

(2) INFORMATION FOR SEQ ID NO: 49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

Met Leu Tyr Ser His Leu Tyr Arg Trp Glu Tyr Thr Ile Pro Phe Leu
1 5 10 15

Leu Leu Leu Ile Met Ala Ser Ser Pro Ser Leu Phe Cys Leu Pro Arg
20 25 30

Ser Leu Lys Ser Val Xaa
35

(2) INFORMATION FOR SEQ ID NO: 50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Met Pro Ala His Phe Leu Phe Ile Leu Ile Phe Phe Val Glu Met Val
1 5 10 15

Ser Cys Cys Val Thr Gln Ala Ser Leu Glu Phe Leu Val Ser Ser Asp
20 25 30

Phe Pro Ala Leu Val Ser Arg Asn Ala Gly Leu Gln Ala Xaa
35 40 45

(2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

Met Phe Leu Val Leu Leu Cys Ile Leu Lys Ser Leu Ile Ile Gly His
1 5 10 15

Val Met Ile Thr Leu Asn Leu Thr Phe Xaa
20 25

(2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

Met Gly Leu Arg Ser Leu Asn Ser Leu Ser Ser Val Met Phe Leu Val
1 5 10 15

His Gly Gly Ala Trp Gly Ala Pro Lys Ala Ser Ala Arg His Leu Cys
20 25 30

Leu Ser Leu Gly Leu His Pro Ser Thr Trp Leu Trp Asn His Asn Arg
35 40 45

Trp Xaa
50

(2) INFORMATION FOR SEQ ID NO: 53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

Gly Glu Thr Phe Ile Pro Ile Ser Leu Ser Xaa
1 5 10

(2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 104 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

Met Lys Gly Trp Gly Trp Leu Ala Leu Leu Leu Gly Ala Leu Leu Gly
1 5 10 15

Thr Ala Trp Ala Arg Arg Ser Gln Asp Leu His Cys Gly Ala Cys Arg
20 25 30

Ala Leu Val Asp Glu Leu Glu Trp Glu Ile Ala Gln Val Asp Pro Lys
35 40 45

Lys Thr Ile Gln Met Gly Ser Phe Arg Ile Asn Pro Asp Gly Ser Xaa
50 55 60

Xaa Val Val Glu Asn Cys Xaa Met Xaa Cys Arg Ala Arg Trp Phe Met
65 70 75 80

Pro Val Val Pro Ala Leu Trp Glu Ala Glu Ala Gly Gly Leu Pro Glu
85 90 95

Val Gly Ser Leu Arg Pro Ala Xaa
100

(2) INFORMATION FOR SEQ ID NO: 55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

Met Thr Leu Thr Leu Leu Leu Ala Leu Ser Pro Lys Pro Pro Val Ser
1 5 10 15

Leu Gly Arg Met Cys Met Gln His Thr His Thr His Thr Lys Thr Lys
20 25 30

Asn Lys Ala Lys Lys Lys Lys Ile Pro Asn Thr Thr Xaa His Asp Cys
35 40 45

Ser Met Thr Pro Asp Ala Gly Arg Thr Gly Cys Xaa
50 55 60

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

Met Ser Leu Thr Leu Leu Met Leu Ile Leu Leu Thr His Asp Leu Arg
1 5 10 15

Asn Lys Phe Ser Leu Lys Cys Ile Phe Cys Arg His Thr Leu Asn Asn
20 25 30

Ile Ile Ser Ser Glu Val Xaa
35

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

Leu Trp Ile Val Ala Val Thr Ser Ser Arg Leu Ser Tyr Leu Gln Glu
1 5 10 15

Asn Met Tyr Met Val Lys Gly His Arg Val Xaa
20 25

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

```
Met Gly Pro Gly Thr Trp Glu Thr Ala Thr Ser Gly Asn Ala Val Gly
 1               5                  10                  15

Glu Gly Gln Ala Leu Gly Ser His Pro Ser Cys Leu Leu Arg Ile Met
            20                  25                  30

Gly Cys Val Gly Arg Val Trp Trp Leu Thr Ser Ile Ile Pro Ala Leu
        35                  40                  45

Trp Glu Ala Lys Val Ser Gly Ser Pro Glu Val Arg Ser Ser Arg Pro
    50                  55                  60

Ala Trp Pro Thr Trp
65
```

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 259 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

```
Met Glu Cys His Leu Lys Thr His Tyr Lys Met Glu Tyr Lys Cys Arg
 1               5                  10                  15

Ile Cys Gln Thr Val Lys Ala Asn Gln Leu Glu Leu Glu Thr His Thr
            20                  25                  30

Arg Glu His Arg Leu Gly Asn His Tyr Lys Cys Asp Gln Cys Gly Tyr
        35                  40                  45

Leu Ser Lys Thr Ala Asn Lys Leu Ile Glu His Val Arg Val His Thr
    50                  55                  60

Gly Glu Arg Pro Phe His Cys Asp Gln Cys Ser Tyr Ser Xaa Lys Arg
65                  70                  75                  80

Lys Asp Asn Leu Asn Leu His Lys Lys Leu Lys His Ala Pro Arg Gln
                85                  90                  95

Thr Phe Ser Cys Glu Glu Cys Leu Phe Lys Thr Thr His Pro Phe Val
            100                 105                 110

Phe Ser Arg His Val Lys Lys His Gln Ser Gly Asp Cys Pro Glu Glu
        115                 120                 125

Asp Lys Lys Gly Leu Cys Pro Ala Pro Lys Glu Pro Ala Gly Pro Gly
    130                 135                 140

Ala Pro Leu Leu Val Val Gly Ser Ser Arg Asn Leu Leu Ser Pro Leu
```

```
145                 150                 155                 160

Ser Val Met Ser Ala Ser Gln Ala Leu Gln Thr Val Ala Leu Ser Ala
                165                 170                 175

Ala His Gly Ser Ser Ser Glu Pro Asn Leu Ala Leu Lys Ala Leu Ala
            180                 185                 190

Phe Asn Gly Ser Pro Leu Arg Phe Asp Lys Tyr Arg Asn Ser Asp Phe
        195                 200                 205

Ala His Leu Ile Pro Leu Thr Met Leu Tyr Pro Lys Asn His Leu Asp
    210                 215                 220

Leu Thr Phe His Pro Pro Arg Pro Gln Thr Ala Pro Pro Ser Ile Pro
225                 230                 235                 240

Ser Pro Lys His Ser Phe Leu Ala Tyr Leu Gly Leu Arg Glu Arg Ala
                245                 250                 255

Glu Thr Val
```

(2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 166 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

```
Met Ser Leu His Val Asp Lys Glu Gln Trp Met Phe Ser Ile Cys Cys
  1               5                  10                  15

Thr Ala Cys Asp Phe Val Thr Met Glu Glu Ala Glu Ile Lys Thr His
             20                  25                  30

Ile Gly Thr Lys His Thr Gly Glu Asp Arg Lys Thr Pro Ser Glu Ser
         35                  40                  45

Asn Ser Pro Ser Ser Ser Ser Leu Ser Ala Leu Ser Asp Ser Ala Asn
     50                  55                  60

Ser Lys Asp Asp Ser Asp Gly Ser Gln Lys Asn Lys Gly Gly Asn Asn
 65                  70                  75                  80

Leu Leu Val Ile Ser Val Met Pro Gly Ser Gln Pro Ser Leu Asn Ser
                 85                  90                  95

Glu Glu Lys Pro Glu Lys Gly Phe Glu Cys Val Phe Cys Asn Phe Val
            100                 105                 110

Cys Lys Thr Lys Asn Met Phe Glu Arg His Leu Gln Ile His Leu Ile
        115                 120                 125

Thr Arg Met Phe Glu Cys Asp Val Cys His Lys Phe Met Lys Thr Pro
    130                 135                 140
```

Glu Gln Leu Leu Glu His Lys Lys Cys His Thr Val Pro Thr Gly Gly
145 150 155 160

Leu Xaa Xaa Gly Gln Trp
165

(2) INFORMATION FOR SEQ ID NO: 61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

Leu Ile Glu His Val Arg Val His Thr Gly Glu Arg Pro Phe His Cys
1 5 10 15

Asp Gln Cys

(2) INFORMATION FOR SEQ ID NO: 62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

Val Asp Pro Lys Lys Thr Ile Gln Met Gly Ser Phe Arg Ile Asn Pro
1 5 10 15

Asp Gly Ser Gln
20

(2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

Tyr Ala Arg Ser Glu Ala His Leu Thr Glu Leu Leu Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 237 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

```
Gly Cys Leu Gly Phe Gln Pro Pro Tyr His Ser Val Pro Ala Trp Glu
 1               5                  10                  15

Arg Ser Thr Arg Gly Gly Asp His Arg Val Glu Leu Tyr Lys Val Leu
            20                  25                  30

Ser Ser Leu Gly Tyr His Val Val Thr Phe Asp Tyr Arg Gly Trp Gly
        35                  40                  45

Asp Ser Val Gly Thr Pro Ser Glu Arg Gly Met Thr Tyr Asp Ala Leu
    50                  55                  60

His Val Phe Asp Trp Ile Lys Ala Arg Ser Gly Asp Asn Pro Val Tyr
65                  70                  75                  80

Ile Trp Gly His Ser Leu Gly Thr Gly Val Ala Thr Asn Leu Val Arg
                85                  90                  95

Arg Leu Cys Glu Arg Glu Thr Pro Pro Asp Ala Leu Ile Leu Glu Ser
            100                 105                 110

Pro Phe Thr Asn Ile Arg Glu Glu Ala Lys Ser His Pro Phe Ser Val
        115                 120                 125

Ile Tyr Arg Tyr Phe Pro Gly Phe Asp Trp Phe Phe Leu Asp Pro Ile
    130                 135                 140

Thr Ser Ser Gly Ile Lys Phe Ala Asn Asp Glu Asn Val Lys His Ile
145                 150                 155                 160

Ser Cys Pro Leu Leu Ile Leu His Ala Glu Asp Asp Pro Val Val Pro
                165                 170                 175

Phe Gln Leu Gly Arg Lys Leu Tyr Ser Ile Ala Ala Pro Ala Arg Ser
            180                 185                 190

Phe Arg Asp Phe Lys Val Gln Phe Val Pro Phe His Ser Asp Leu Gly
        195                 200                 205

Tyr Arg His Lys Tyr Ile Tyr Lys Ser Pro Glu Leu Pro Arg Ile Leu
    210                 215                 220

Arg Glu Phe Leu Gly Lys Ser Glu Pro Glu His Gln His
225                 230                 235
```

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

```
Tyr Arg Gly Trp Gly Asp Ser Val Gly Thr Pro Ser Glu Arg Gly Met
 1               5                  10                  15

Thr Tyr Asp
```

(2) INFORMATION FOR SEQ ID NO: 66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

Ala Leu Ile Leu Glu Ser Pro Phe Thr Asn Ile
1 5 10

(2) INFORMATION FOR SEQ ID NO: 67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 442 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

Leu Asp Ala Val Leu Glu Tyr Leu Pro Asn Pro Ser Glu Val Gln Asn
1 5 10 15

Tyr Ala Ile Leu Asn Lys Glu Asp Asp Ser Lys Glu Lys Thr Lys Ile
20 25 30

Leu Met Asn Ser Ser Arg Asp Asn Ser His Pro Phe Val Gly Leu Ala
35 40 45

Phe Lys Leu Glu Val Gly Arg Phe Gly Gln Leu Thr Tyr Val Arg Ser
50 55 60

Tyr Gln Gly Glu Leu Lys Lys Gly Asp Thr Ile Tyr Asn Thr Arg Thr
65 70 75 80

Arg Lys Lys Val Arg Leu Gln Arg Leu Ala Arg Met His Ala Asp Met
85 90 95

Met Glu Asp Val Glu Glu Val Tyr Ala Gly Asp Ile Cys Ala Leu Phe
100 105 110

Gly Ile Asp Cys Ala Ser Gly Asp Thr Phe Thr Asp Lys Ala Asn Ser
115 120 125

Gly Leu Ser Met Glu Ser Ile His Val Pro Asp Pro Val Ile Ser Ile
130 135 140

Ala Met Lys Pro Ser Asn Lys Asn Asp Leu Glu Lys Phe Ser Lys Gly
145 150 155 160

Ile Gly Arg Phe Thr Arg Glu Asp Pro Thr Phe Lys Val Tyr Phe Asp
165 170 175

Thr Glu Asn Lys Glu Thr Val Ile Ser Gly Met Gly Glu Leu His Leu
180 185 190

```
Glu Ile Tyr Ala Gln Arg Leu Glu Arg Glu Tyr Gly Cys Pro Cys Ile
        195                 200                 205

Thr Gly Lys Pro Lys Val Ala Phe Arg Glu Thr Ile Thr Ala Pro Val
    210                 215                 220

Pro Phe Asp Phe Thr His Lys Lys Gln Ser Gly Gly Ala Gly Gln Tyr
225                 230                 235                 240

Gly Lys Val Ile Gly Val Leu Glu Pro Leu Asp Pro Glu Asp Tyr Thr
                245                 250                 255

Lys Leu Glu Phe Ser Asp Glu Thr Phe Gly Ser Asn Ile Pro Lys Gln
            260                 265                 270

Phe Val Pro Ala Val Glu Lys Gly Phe Leu Asp Ala Cys Glu Lys Gly
        275                 280                 285

Pro Leu Ser Gly His Lys Leu Ser Gly Leu Arg Phe Val Leu Gln Asp
    290                 295                 300

Gly Ala His His Met Val Asp Ser Asn Glu Ile Ser Phe Ile Arg Ala
305                 310                 315                 320

Gly Glu Gly Ala Leu Lys Gln Ala Leu Ala Asn Ala Thr Leu Cys Ile
                325                 330                 335

Leu Glu Pro Ile Met Ala Val Glu Val Val Ala Pro Asn Glu Phe Gln
            340                 345                 350

Gly Gln Val Ile Ala Gly Ile Asn Arg Arg His Gly Val Ile Thr Gly
        355                 360                 365

Gln Asp Gly Val Glu Asp Tyr Phe Thr Leu Tyr Ala Asp Val Pro Leu
    370                 375                 380

Asn Asp Met Phe Gly Tyr Ser Thr Glu Leu Arg Ser Cys Thr Glu Gly
385                 390                 395                 400

Lys Gly Glu Tyr Thr Met Glu Tyr Ser Arg Tyr Gln Pro Cys Leu Pro
                405                 410                 415

Ser Thr Gln Glu Asp Val Ile Asn Lys Tyr Leu Glu Ala Thr Gly Gln
            420                 425                 430

Leu Pro Val Lys Lys Gly Lys Ala Lys Asn
        435                 440
```

(2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

```
Ser His Pro Phe Val Gly Leu Ala Phe Lys Leu Glu
```

1 5 10

(2) INFORMATION FOR SEQ ID NO: 69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

Arg Met His Ala Asp Met Met Glu Asp Val Glu Glu Val Tyr Ala Gly
1 5 10 15

Asp Ile Cys Ala Leu Phe Gly Ile Asp Cys Ala Ser Gly Asp
20 25 30

(2) INFORMATION FOR SEQ ID NO: 70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

Leu Ser Met Glu Ser Ile His Val Pro Asp Pro Val Ile Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

Ala Met Lys Pro Ser Asn Lys Asn Asp Leu Glu Lys Phe Ser Lys Gly
1 5 10 15

Ile

(2) INFORMATION FOR SEQ ID NO: 72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

Arg Phe Thr Arg Glu Asp Pro Thr Phe Lys Val
1 5 10

(2) INFORMATION FOR SEQ ID NO: 73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

Phe Val Leu Gln Asp Gly Ala His His Met Val Asp Ser Asn Glu Ile
1 5 10 15

Ser Phe Ile Arg Ala Gly Glu Gly Ala Leu Lys Gln Ala Leu Ala
20 25 30

(2) INFORMATION FOR SEQ ID NO: 74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Glu Asp Tyr Phe Thr Leu Tyr Ala Asp Val Pro Leu Asn Asp Met Phe
1 5 10 15

Gly Tyr Ser Thr Glu Leu Arg Ser Cys Thr Glu Gly Lys Gly Glu Tyr
20 25 30

Thr Met Glu Tyr
35

(2) INFORMATION FOR SEQ ID NO: 75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

Gly Gln Leu Pro Val Lys Lys Gly Lys Ala Lys Asn
1 5 10

(2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 294 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

Met Gly Ser Thr Val Cys Thr Asp Glu Arg Xaa Met Ala Glu Leu Ala
1 5 10 15

```
Lys Glu Leu Pro Gln Val Ser Phe Val Lys Leu Glu Ala Glu Gly Val
            20                  25                  30

Pro Glu Val Ser Glu Lys Tyr Glu Ile Ser Ser Val Pro Thr Phe Leu
        35                  40                  45

Phe Phe Lys Asn Ser Gln Lys Ile Asp Arg Leu Asp Gly Ala His Ala
    50                  55                  60

Pro Glu Leu Thr Lys Lys Val Gln Arg His Ala Ser Ser Gly Ser Phe
65                  70                  75                  80

Leu Pro Ser Ala Asn Glu His Leu Lys Glu Asp Leu Asn Leu Arg Leu
                85                  90                  95

Lys Lys Leu Thr His Ala Ala Pro Cys Met Leu Phe Met Lys Gly Thr
            100                 105                 110

Pro Gln Glu Pro Arg Cys Gly Phe Ser Lys Gln Met Val Glu Ile Leu
        115                 120                 125

His Lys His Asn Ile Gln Phe Ser Ser Phe Asp Ile Phe Ser Asp Glu
    130                 135                 140

Glu Val Arg Gln Gly Leu Lys Ala Tyr Ser Ser Trp Pro Thr Tyr Pro
145                 150                 155                 160

Gln Leu Tyr Val Ser Gly Glu Leu Ile Gly Gly Leu Asp Ile Ile Lys
                165                 170                 175

Glu Leu Glu Ala Ser Glu Glu Leu Asp Thr Ile Cys Pro Lys Ala Pro
            180                 185                 190

Lys Leu Glu Glu Arg Leu Lys Val Leu Thr Asn Lys Ala Ser Val Met
        195                 200                 205

Leu Phe Met Lys Gly Asn Lys Gln Glu Ala Lys Cys Gly Phe Ser Lys
    210                 215                 220

Gln Ile Leu Glu Ile Leu Asn Ser Thr Gly Val Glu Tyr Glu Thr Phe
225                 230                 235                 240

Asp Ile Leu Glu Asp Glu Glu Val Arg Gln Gly Leu Lys Ala Tyr Ser
                245                 250                 255

Asn Trp Pro Thr Tyr Pro Gln Leu Tyr Val Lys Gly Glu Leu Val Gly
            260                 265                 270

Gly Leu Asp Ile Val Lys Glu Leu Lys Glu Asn Gly Glu Leu Leu Pro
        275                 280                 285

Ile Leu Arg Gly Glu Asn
    290
```

(2) INFORMATION FOR SEQ ID NO: 77:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

Met Leu Phe Met Lys Gly Thr Pro Gln Glu Pro Arg Cys Gly Phe Ser
1 5 10 15

Lys Gln Met Val Glu Ile Leu
20

(2) INFORMATION FOR SEQ ID NO: 78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

Trp Pro Thr Tyr Pro Gln Leu Tyr Val Ser Gly Glu Leu Ile Gly Gly
1 5 10 15

Leu Asp Ile Ile Lys Glu
20

(2) INFORMATION FOR SEQ ID NO: 79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

Phe Lys His Arg Gly Leu Glu Tyr Gly Arg Phe Leu Arg Xaa Trp Glu
1 5 10 15

Leu Lys Pro Glu Phe Xaa Lys Gly Phe Arg Thr Asp Gly Arg Ala Gly
20 25 30

Xaa Trp Val Xaa Gly Asp Phe Gly Lys Arg Phe Phe Arg Pro Gly Glu
35 40 45

Val Ala Asp Ser Cys Asn Pro Ser Thr Phe Gly Xaa Arg Gly Trp Gln
50 55 60

Ile Thr Cys Arg Pro Gly Val
65 70

(2) INFORMATION FOR SEQ ID NO: 80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

Gly Asp Phe Gly Lys Arg Phe Phe Arg Pro Gly Glu Val Ala Asp Ser
1 5 10 15

Cys Asn Pro Ser Thr Phe Gly
20

(2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

Met Gly Gly Gln Val Xaa Gly Ser Xaa Xaa Ile Leu Glu Lys Asp Phe
1 5 10 15

Ser Gly Gln Val Arg Trp Leu Ile Pro Val Ile Pro Ala Leu Leu Glu
20 25 30

Xaa Glu Ala Gly Arg Ser Leu Val Gly Gln Glu Phe Glu Thr Ser Leu
35 40 45

Gly Asn Met Ala Lys Pro Cys Leu Tyr Lys Asn Tyr Lys Ile Ser Ala
50 55 60

Arg Ser Gly Gly Leu Cys Leu
65 70

(2) INFORMATION FOR SEQ ID NO: 82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

Ile Leu Glu Lys Asp Phe Ser Gly Gln Val Arg Trp Leu Ile Pro Val
1 5 10 15

Ile Pro Ala Leu Leu Glu
20

(2) INFORMATION FOR SEQ ID NO: 83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

Glu Ala Gly Arg Ser Leu Val Gly Gln Glu Phe Glu Thr Ser Leu Gly

```
 1               5                  10                  15

Asn Met Ala Lys Pro Cys Leu Tyr Lys Asn Tyr Lys Ile Ser Ala Arg
            20                  25                  30

Ser Gly Gly Leu Cys Leu
        35
```

(2) INFORMATION FOR SEQ ID NO: 84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

```
Met Thr Val Gly Pro Ala Ser Ala Leu Phe Pro Cys Gln Thr Pro Xaa
 1               5                  10                  15

Phe Pro Trp Thr Glu Trp Asn Xaa Trp Glu Phe Thr Ala His Val Leu
            20                  25                  30

Ser Gln Lys Phe Glu Lys Glu Leu Ser Lys Val Arg Glu Tyr Val Gln
        35                  40                  45

Leu Ile Ser Val Tyr Glu Lys Lys Leu Leu Asn Leu Thr Val Arg Ile
    50                  55                  60

Asp Ile Met Glu Lys Asp Thr Ile Ser Tyr Xaa Glu Leu Asp Phe Glu
65                  70                  75                  80

Leu Ile Lys Val Glu Val Lys Glu Met Glu Lys Leu Val Ile Gln Leu
                85                  90                  95

Lys Glu Pro Phe Gly Gly Ser Ser Glu Ile Val Gly Pro Ala Gly Gly
            100                 105                 110

Gly Asp Lys Lys Tyr Asp Ser Leu Gly Arg Glu Ala
        115                 120
```

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 318 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

```
Met Thr Leu Leu Val Glu Lys Leu Glu Thr Leu Asp Lys Asn Xaa Val
 1               5                  10                  15

Leu Ala Ile Arg Arg Glu Xaa Val Ala Leu Lys Thr Lys Leu Lys Glu
            20                  25                  30

Cys Glu Ala Ser Lys Asp Gln Asn Thr Pro Val Val His Pro Pro Pro
        35                  40                  45
```

```
Thr Pro Gly Ser Cys Gly His Gly Gly Val Val Xaa Ile Ser Lys Pro
     50                  55                  60

Ser Val Val Gln Leu Asn Trp Arg Gly Phe Ser Tyr Leu Tyr Gly Ala
 65                  70                  75                  80

Trp Gly Arg Asp Tyr Ser Pro Gln His Pro Asn Lys Gly Leu Tyr Trp
                 85                  90                  95

Val Ala Pro Leu Asn Thr Asp Gly Arg Leu Leu Glu Tyr Tyr Arg Leu
            100                 105                 110

Tyr Asn Thr Leu Asp Asp Leu Leu Leu Tyr Ile Asn Ala Arg Glu Leu
        115                 120                 125

Arg Ile Thr Tyr Gly Gln Gly Ser Gly Thr Ala Val Tyr Asn Asn Asn
    130                 135                 140

Met Tyr Val Asn Met Tyr Asn Thr Gly Asn Ile Ala Arg Val Asn Leu
145                 150                 155                 160

Thr Thr Asn Thr Ile Ala Val Thr Gln Thr Leu Pro Asn Ala Ala Tyr
                165                 170                 175

Asn Asn Arg Phe Xaa Tyr Ala Asn Val Ala Trp Gln Asp Ile Asp Phe
            180                 185                 190

Xaa Val Asp Glu Asn Gly Leu Trp Val Ile Tyr Ser Thr Glu Ala Ser
        195                 200                 205

Thr Gly Asn Met Val Ile Ser Lys Leu Asn Asp Thr Thr Leu Gln Val
    210                 215                 220

Leu Asn Thr Trp Tyr Thr Xaa Gln Tyr Lys Pro Ser Ala Ser Asn Ala
225                 230                 235                 240

Phe Met Val Cys Gly Val Leu Tyr Ala Thr Arg Thr Met Asn Thr Arg
                245                 250                 255

Thr Glu Glu Ile Phe Tyr Tyr Tyr Asp Thr Asn Thr Gly Lys Glu Gly
            260                 265                 270

Lys Leu Asp Ile Val Met His Lys Met Gln Glu Lys Val Gln Ser Ile
        275                 280                 285

Asn Tyr Asn Pro Phe Asp Gln Lys Leu Tyr Val Tyr Asn Asp Gly Tyr
    290                 295                 300

Leu Leu Asn Tyr Asp Leu Ser Val Leu Gln Lys Pro Gln Cys
305                 310                 315
```

(2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

Leu Glu Thr Leu Asp Lys Asn Xaa Val Leu Ala Ile Arg Arg Glu Xaa
1 5 10 15

Val Ala Leu Lys Thr Lys Leu Lys Glu Cys Glu
20 25

(2) INFORMATION FOR SEQ ID NO: 87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

Tyr Trp Val Ala Pro Leu Asn Thr Asp Gly Arg Leu Leu Glu
1 5 10

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

Ala Ser Asn Ala Phe Met Val Cys Gly Val Leu Tyr
1 5 10

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

Thr Gly Lys Glu Gly Lys Leu Asp Ile Val Met
1 5 10

(2) INFORMATION FOR SEQ ID NO: 90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

Met Ser Arg Leu Leu Ala Lys Ala Lys Asp Phe Arg Tyr Asn Leu Ser
1 5 10 15

```
Glu Val Leu Gln Gly Lys Leu Gly Ile Tyr Asp Ala Asp Gly Asp Gly
            20                  25                  30

Asp Phe Asp Val Asp Asp Ala Lys Val Leu Leu Gly Leu Thr Lys Asp
        35                  40                  45

Gly Ser Asn Glu Asn Ile Asp Ser Leu Glu Glu Val Leu Asn Ile Leu
    50                  55                  60

Ala Glu Glu Ser Ser Asp Trp Phe Tyr Gly Phe Leu Ser Phe Leu Tyr
65                  70                  75                  80

Asp Ile Met Thr Pro Phe Glu Met Leu Glu Glu Glu Glu Glu Glu Ser
                85                  90                  95

Glu Thr Ala Asp Gly Val Asp Gly Thr Ser Gln Asn Glu Gly Val Gln
            100                 105                 110

Gly Lys Thr Cys Val Ile Leu Asp Leu His Asn Gln
        115                 120
```

(2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

```
Thr Ser Ala Gly Ser Ser Ser Pro Gly Thr Arg Glu Arg Asp Lys Ala
 1               5                  10                  15

Trp Arg Thr Gln Gln Trp Glu Glu Arg Arg Thr Leu Arg Asn Phe Ile
            20                  25                  30

Leu His Val Val Tyr Gly Asp Cys Ile Ala Gly Arg Leu Asp Ile Cys
        35                  40                  45

Thr Cys Arg Leu Val
    50
```

(2) INFORMATION FOR SEQ ID NO: 92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

```
Arg Val Arg Ala Ala Ala Ala Pro Ala Arg Gly Arg Glu Thr Lys His
 1               5                  10                  15

Gly Gly His Asn Asn
            20
```

(2) INFORMATION FOR SEQ ID NO: 93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

```
Ser Phe Phe Thr Trp Phe Met Val Ile Ala Leu Leu Gly Val Trp Thr
 1               5                  10                  15

Ser Val
```

(2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

```
Trp Cys Gln Arg Val Gln Asp Leu Ser Ala Arg Val Arg Gly Glu Gln
 1               5                  10                  15

Cys Cys Ala Val Gly Arg Asn Leu Thr Ile Thr Gln Ser Pro Arg Gln
            20                  25                  30

Arg Val Gln Asp Leu Ser Thr Gly Val Arg Gly Glu Gln Arg Cys Pro
        35                  40                  45

Ala Gly Arg Ser Leu Thr Ile Thr Gln Ser Pro His Arg His Pro Val
    50                  55                  60

Ser Ser Pro Glu Gly Pro Gly Pro Gln Cys Arg Gly Ala Arg Arg Ala
65                  70                  75                  80

Val Leu Ser Ser Gly Glu Glu Pro His His His Ser Val Ser Ser Pro
                85                  90                  95

Ala His Phe Phe Ser Met Ser Arg Phe Ala Pro Pro Leu Val Phe Val
            100                 105                 110

Phe Leu Lys Glu Asp Phe Glu Lys Arg Trp
        115                 120
```

(2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 156 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

```
Asn Gln Leu Thr Phe Ile Trp Lys Lys Pro His Phe Thr Val Val Cys
```

```
 1               5                  10                  15

His Phe Asp Gly Val Arg Gly Ser Arg Thr Ser Val Pro Gly Cys Glu
            20                  25                  30

Glu Ser Ser Ala Val Gln Trp Gly Gly Thr Ser Pro Ser Pro Ser Leu
        35                  40                  45

Leu Ala Arg Gly Ser Arg Thr Ser Val Pro Gly Cys Glu Glu Ser Ser
    50                  55                  60

Ala Val Gln Arg Gly Gly Val Ser Pro Ser Pro Ser Leu Leu Thr Val
65                  70                  75                  80

Thr Gln Ser Pro Arg Gln Arg Val Gln Asp Leu Ser Ala Gly Val Arg
                85                  90                  95

Gly Glu Gln Cys Cys Pro Ala Gly Arg Asn Leu Thr Ile Thr Gln Ser
            100                 105                 110

Pro His Gln His Thr Phe Ser Pro Cys Leu Val Leu Leu Leu Leu Trp
        115                 120                 125

Tyr Leu Tyr Phe Leu Lys Arg Ile Leu Lys Arg Asp Gly Glu Val Gly
    130                 135                 140

Ile Leu Gly Arg Arg Asp Gln Leu Phe Pro Gln Asp
145                 150                 155
```

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 129 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

```
Leu Ser Phe Gly Lys Ser Pro Thr Ser Leu Trp Ser Val Thr Leu Met
 1               5                  10                  15

Val Ser Glu Gly Pro Gly Pro Gln Cys Gln Gly Ala Arg Arg Ala Val
            20                  25                  30

Leu Cys Ser Gly Glu Glu Pro His His His Pro Val Ser Ser Pro Glu
        35                  40                  45

Gly Pro Gly Pro Gln Tyr Arg Gly Ala Arg Arg Ala Ala Leu Ser Ser
    50                  55                  60

Gly Glu Glu Ser His His His Pro Val Ser Ser Pro Ser Pro Ser Leu
65                  70                  75                  80

Leu Ala Arg Gly Ser Arg Thr Ser Val Pro Gly Cys Glu Glu Ser Ser
                85                  90                  95

Ala Val Gln Arg Gly Gly Thr Ser Pro Ser Leu Ser Leu Leu Thr Ser
            100                 105                 110
```

```
Thr Leu Phe Leu His Val Ser Phe Cys Ser Ser Ser Gly Ile Cys Ile
        115                 120                 125

Ser
```

(2) INFORMATION FOR SEQ ID NO: 97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 114 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

```
Met Val Ser Glu Gly Pro Gly Pro Gln Cys Gln Gly Ala Arg Arg Ala
  1               5                  10                  15

Val Leu Cys Ser Gly Glu Glu Pro His His His Pro Val Ser Ser Pro
             20                  25                  30

Glu Gly Pro Gly Pro Gln Tyr Arg Gly Ala Arg Arg Ala Ala Leu Ser
         35                  40                  45

Ser Gly Glu Glu Ser His His His Pro Val Ser Ser Pro Ser Pro Ser
     50                  55                  60

Leu Leu Ala Arg Gly Ser Arg Thr Ser Val Pro Gly Cys Glu Glu Ser
 65                  70                  75                  80

Ser Ala Val Gln Arg Gly Gly Thr Ser Pro Ser Leu Ser Leu Leu Thr
                 85                  90                  95

Ser Thr Leu Phe Leu His Val Ser Phe Cys Ser Ser Ser Gly Ile Cys
            100                 105                 110

Ile Ser
```

(2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 212 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

```
Gly Leu Cys Thr Glu Val Ala Phe Ala Ala Ser Leu Arg Gly Pro Ser
  1               5                  10                  15

Ala His Ile Ile Ser Asp Pro Gln Thr Thr Leu Gln Arg Gly Gly Arg
             20                  25                  30

Cys Cys Lys Leu His Ser Ser Pro Asn Trp His His Pro Ala Ser Trp
         35                  40                  45

Asp Ser Asp Gln Gly Cys Gln Thr Pro Glu Pro Val Val Leu Ser Leu
```

```
     50                  55                  60

His Leu Ser Ala Arg Pro Pro Pro Trp Ser Gly Phe Leu Ser Phe Leu
 65                  70                  75                  80

Leu Gln Val Ser Phe Ser Leu Cys Tyr His Leu Cys Ser Glu Gln Leu
                85                  90                  95

Leu Thr Thr Gln Arg Val Ser Cys Ala His Ile Tyr Ser Ala Leu Asp
            100                 105                 110

Pro Thr Ala Arg Lys Ile Asn Leu Ala Lys Phe Thr Leu Gly Lys Cys
        115                 120                 125

Ser Thr Leu Ile Val Thr Asp Leu Ala Ala Arg Gly Leu Asp Ile Pro
    130                 135                 140

Leu Leu Asp Asn Val Ile Asn Tyr Ser Phe Pro Ala Lys Gly Lys Leu
145                 150                 155                 160

Phe Leu His Arg Val Gly Lys Gln Pro Val Ala Gly Pro Gly Ala Gly
                165                 170                 175

Arg Gly Ala Gly Ser Trp Gln Lys Pro Arg Val Gln Gly Leu Thr Leu
            180                 185                 190

Asp Thr Ala His Gly Val Ala Val Gly Leu Val Leu Glu Thr Glu Pro
        195                 200                 205

Arg Tyr Ile Ala
    210
```

(2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

```
Val Thr Asp Leu Ala Ala Arg Gly Leu Asp Ile Pro Leu Leu Asp Asn
 1               5                  10                  15

Val Ile Asn Tyr Ser Phe
            20
```

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

```
Gly Ile Glu Lys Phe Gly Asn Leu Pro Lys Val Thr Gln Leu Val Cys
 1               5                  10                  15
```

Ser Arg Ile Arg Ile Arg Leu Val His
20 25

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

Lys Ser Leu Val Thr Cys Pro Arg Ser His Ser Leu Phe Val Ala Glu
1 5 10 15

Ser Gly

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

Val Phe His Val Glu Thr Leu Phe Ser Ala Leu Tyr Ile Leu Thr His
1 5 10 15

Val Ile Leu Ile Ile Arg His Lys Glu Gly Ala Val Ile Arg Thr Asp
20 25 30

Glu Glu Asn Glu Ala
35

(2) INFORMATION FOR SEQ ID NO: 103:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 69 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

Thr Phe Gln Phe Cys His Thr His Gln Pro Cys Thr Cys Pro Ser His
1 5 10 15

His Ser Gly Tyr Lys Ser Ile Ser Leu Trp Phe Trp Leu Cys Pro Asn
20 25 30

Asp Cys Glu Ala Glu His Leu Phe Lys Cys Glu Leu Ala Ile Tyr Ile
35 40 45

Pro Ser Leu Glu Asn Cys Leu Phe Lys Pro Phe Ala Pro Phe Tyr Ile
50 55 60

Glu Leu Ser Ile Phe
65

(2) INFORMATION FOR SEQ ID NO: 104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 90 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

Leu Tyr Tyr Phe Ile Phe Pro Pro Ala Val Asn Lys His Ser Asn Phe
1 5 10 15

Ala Ile Leu Thr Asn Leu Val Leu Val Gln Ala Ile Ile Val Gly Ile
20 25 30

Lys Val Phe Pro Cys Gly Ser Gly Tyr Ala Leu Met Thr Val Arg Leu
35 40 45

Asn Ile Phe Ser Ser Val Asn Trp Pro Phe Ile Tyr Leu Leu Trp Arg
50 55 60

Thr Val Phe Ser Asn Pro Leu Leu Leu Phe Thr Leu Ser Tyr Pro Ser
65 70 75 80

Phe Asn Cys Trp Val Val Tyr Cys Leu Ile
85 90

(2) INFORMATION FOR SEQ ID NO: 105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

His Gln Ala Pro Thr Gln Ser Gln Leu Gly Asn Gln Ser His Pro Pro
1 5 10 15

Trp Leu Cys Trp Gly Gly Pro Ala Ile Cys Pro Trp Ser Arg Arg Glu
20 25 30

Arg Gly Val Ser Pro Arg Pro Gly Ala Gly Lys Glu Cys Val Pro Gln
35 40 45

Leu Ser Ala Leu Leu Ile Leu Ile Met Glu Lys Pro Leu Phe Leu Ser
50 55 60

Pro Phe Pro Glu Leu Val Phe Cys Cys Phe Cys Phe Ile Leu Phe Trp
65 70 75 80

Gly Asp Ser Phe Leu Leu Phe Asn Leu Glu Ser Pro Val Pro Leu Gly
85 90 95

```
Cys Arg Gln Phe Leu Pro Gly Pro Ser Arg Asn Pro His Ser Pro Ser
            100                 105                 110

Pro Leu Leu Arg Tyr Leu Gln Glu Ala Ala Asn Leu Val His Ser Asp
        115                 120                 125

Lys Pro Pro Thr Gln Ile Ser Leu Leu Pro Leu Cys Pro Lys Ser His
    130                 135                 140

His
145
```

(2) INFORMATION FOR SEQ ID NO: 106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 89 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

```
Met Glu Lys Pro Leu Phe Leu Ser Pro Phe Pro Glu Leu Val Phe Cys
  1               5                  10                  15

Cys Phe Cys Phe Ile Leu Phe Trp Gly Asp Ser Phe Leu Leu Phe Asn
             20                  25                  30

Leu Glu Ser Pro Val Pro Leu Gly Cys Arg Gln Phe Leu Pro Gly Pro
         35                  40                  45

Ser Arg Asn Pro His Ser Pro Ser Pro Leu Leu Arg Tyr Leu Gln Glu
     50                  55                  60

Ala Ala Asn Leu Val His Ser Asp Lys Pro Pro Thr Gln Ile Ser Leu
 65                  70                  75                  80

Leu Pro Leu Cys Pro Lys Ser His His
                85
```

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 33, line N/A.

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution: American Type Culture Collection

Address of depositary institution (*including postal code and country*):

10801 University Boulevard
Manassas, Virginia 20110-2209
United States of America

Date of deposit	Accession Number
June 12, 1997	209118

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later (*specify the general nature of the indications. e.g., "Accession Number of Deposit"*)

For receiving Office use only

☒ This sheet was received with the international application

Authorized officer

For International Bureau use only

☐ This sheet was received by the International Bureau on:

Authorized officer

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.